



**A University of Sussex PhD thesis**

Available online via Sussex Research Online:

<http://sro.sussex.ac.uk/>

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details

# **Radical Sensorimotor Enactivism**

**Adrian Downey**

PhD Philosophy

University of Sussex

September, 2016

University of Sussex

Adrian Downey

PhD Philosophy

Radical Sensorimotor Enactivism

### **Abstract**

In this thesis I develop a novel approach to conscious perception, which I label “radical sensorimotor enactivism” (RSE). In chapter one, I explain how the development of RSE is guided by the tenets of *activity* and *knowledge-how*. In chapter two, I outline and explain RSE. Throughout the thesis, I will pit RSE against cognitivist accounts of conscious perception and argue that RSE is to be preferred.

In chapters three and four, I highlight two problems facing cognitivist accounts of conscious perception which RSE avoids. I argue that cognitivist accounts of conscious perception face the ‘hard problem of perceptual consciousness’, whilst RSE can provide a phenomenologically plausible deflation of this problem. I next explain why cognitivist accounts are incapable of providing a satisfactory explanation of split-brain syndrome. Then, I argue that RSE can provide a parsimonious explanation of this syndrome.

Theories predicated on *activity* and *knowledge-how* are often rejected for being incapable of accounting for the brain’s role in conscious perception. In chapter five, I argue that RSE can account for the brain’s role by adopting a non-representational version of predictive processing (PP). Moreover, I argue that the resultant account improves upon cognitivist alternatives. Then, in chapter six, I argue that even representational explanations of PP can be subsumed within RSE by accepting fictionalism about their representational posits. Consequently, I conclude that RSE cannot be objected to for failing to account for the brain’s role in conscious perception.

Finally, in chapter seven, I discuss ‘non-veridical’ experiences. Accounts like RSE are often rejected because it is thought they are incapable of explaining the existence of these phenomena. I explain how the existence of such phenomena is wholly compatible with the truth of RSE. Thus, I conclude that RSE should not be rejected solely on the basis that non-veridical experiences exist.

## Acknowledgements

This thesis is the product of three years of (hard) work, which was carried out in a variety of locations. I would like to thank all the people who I have encountered in academic settings throughout this time— whether at the universities of Sussex and British Columbia, academic conferences, or elsewhere. Particular thanks go to all the graduate students at Sussex and UBC (whom I will not list, for fear of missing out a name), the MLEC crew and associates, Benjamin Jarvis, J. Adam Carter, Joe Morrison, Mahon O’Brien, and Tony Booth. I would especially like to thank my academic supervisors— Chris Mole (who provided extremely helpful and stimulating feedback and discussion on this thesis), and Sarah Sawyer (whose advice and training was not only essential to the content of this thesis, but was indispensable for my general development as a philosopher).

Thanks of a different sort go to UBC Judo, Sussex IMA, Chum Sut MMA, and St. John’s CLG (for providing sporting distraction) and the various cafes I spent an inordinate amount of time in (for providing the coffee). I would also like to thank my family and friends for their help and support. Finally, I would like to thank the University of Sussex for awarding me an AHRC scholarship, and the AHRC for honouring it. Without the financial support, this thesis would never have been written.

## Contents

Introduction .....	8
Chapter One: A Rylean Cognitive Science.....	14
§1: Cartesian Cognitive Science .....	15
§2: Rylean Cognitive Science .....	17
§3: Conclusion.....	25
Chapter Two: Radical Sensorimotor Enactivism .....	27
§1: What is Sensorimotor Enactivism? .....	28
§2: Radicalising Sensorimotor Enactivism .....	30
§3: ‘Radical’ Sensorimotor Enactivism.....	44
§4: Conclusion.....	46
Chapter Three: Cognitivism, Radical Sensorimotor Enactivism, and the ‘Hard Problem of Perceptual Consciousness’ .....	48
§1: The Sub-Personal, Personal, and Conscious Levels of Explanation .....	49
§2: Cognitivism and the ‘Hard Problem of Perceptual Consciousness’ .....	65
§3: Dissolving the ‘Hard Problem of Perceptual Consciousness’ .....	76
§4: Conclusion.....	88
Chapter Four: Split-Brain Syndrome .....	90
Part One .....	91
§1: What is a split-brain? .....	91
§2: The Interpreter Model .....	96

§3: Two-streams Model .....	104
§4: Partial Unity .....	108
§5: Switch Model .....	112
§6: Conclusion of Part One .....	117
Part Two: RSEcuing Split-Brain Syndrome .....	120
§1: Interpretation Five- Experimental Aberration Model.....	120
§2: The Nomological Possibility of an Externalist Account of Split-brain Syndrome.....	123
§3: From Possibility to Actuality— the Empirical Evidence .....	127
§4: Split-Brain Syndrome and Sensorimotor Skill .....	131
§5: RSE and the Experimental Aberration Interpretation.....	142
§6: Conclusion.....	150
§7: Postscript— Tye and Marks’ Argument .....	151
Interlude: The Benefits of Activity and Knowledge-How.....	153
Chapter Five: RSE and the Brain .....	155
§1: What is Binocular Rivalry and Why is it Pursued as a Cognitivist Research Programme?.....	156
§2: PP and Binocular Rivalry .....	158
§3: Non-Representational PP.....	164
§4: RSE & Rivalry .....	179
§5: Conclusion.....	186
Chapter Six: Fictionalism about the Representational Posits of PP.....	188
§1: Motivating Fictionalism.....	189
§2: Models As Make-Believe and Predictive Processing .....	191

§3: Benefits of Fictionalism.....	204
§4: Conclusion.....	213
Chapter Seven: A Note On ‘Non-veridical’ Experience .....	216
§1: Does Experience Supervene Upon the Brain? .....	218
§2: Hallucination .....	224
§3: Illusion .....	233
§4: Non-veridical Experience and Representation .....	241
§5: Conclusion.....	243
Conclusion of Thesis.....	245
Bibliography .....	250

## Introduction

In this thesis I develop a framework for the scientific study of conscious perception that is both phenomenologically plausible and empirically fruitful. I set up my theory in direct opposition to the dominant *cognitivist* paradigm in cognitive science, and argue that it is to be preferred over its cognitivist rivals because it provides better explanations of conscious perception at both the personal level of explanation and the sub-personal level of empirical implementation.

In chapter one I introduce the two tenets (which I derive from Ryle [1949/2000]) that serve to constrain and guide the development of my theory:

*Activity* — mentality is an activity which an organism performs.

*Knowledge-How* — the ability to engage in mental activity is predicated on an organism's possession of the requisite knowledge-how, and should not be explained in terms of representation.

Then, I explain that *cognitivist* accounts can be defined as accepting the negation of both tenets. Throughout the thesis I will argue that my own account provides better explanations of conscious perception than cognitivist ones, and so will conclude that it should be preferred as an approach to the scientific study of conscious perception on this basis.

In chapter two I introduce my own theory of conscious perception, which I label "*radical sensorimotor enactivism*" (RSE). RSE is based on the *sensorimotor enactive* (SE) approach to conscious perception [O'Regan & Noë, 2001; Noë, 2004; O'Regan, 2011]. SE takes perception to be constituted by a direct relation between the organism and its environment which is enabled by the organism's possession of *sensorimotor knowledge*.



*Conscious* perception occurs if the organism attends to the occurrent exercise of sensorimotor knowledge, and so on SE attention is necessary for consciousness.

Although SE fulfils the criterion of *activity* it does not fulfil the criterion of *knowledge-how*, because extant SE accounts of sensorimotor knowledge and attention appear to require representation. I argue that sensorimotor knowledge should be understood entirely in terms of causal mediation (at the sub-personal level) and behavioural dispositions (at the personal level), and that a non-representational and adverbial theory of attention should be applied to SE, in order to fulfil the *knowledge-how* constraint. Having modified SE in this manner, I therefore arrive at my own non-representational (or radical) version of the view, and so at *radical* sensorimotor enactivism.

RSE understands conscious perception to be an activity an organism performs which is predicated on its possessing the relevant knowledge-how to perceive. Perception is taken to be a type of behavioural relation to the environment, and conscious perception is understood to occur when the organism engages in this perceptual behaviour ‘attentively’. As such, both perception and conscious perception are explained in terms of behaviour. RSE ascribes to the brain the role of controlling behaviour, and so brains ‘give rise’ to perception and consciousness because they control behaviour. Having stated the theory in chapter two, the rest of the thesis is dedicated to arguing that RSE is to be preferred over its cognitivist rivals.

In chapter three I distinguish between the *sub-personal*, *personal*, and *conscious* levels of explanation. I then introduce and explain the cognitivist theory of conscious perception known as *predictive processing* (PP), and explain how this theory proposes to account for these levels of explanation. I argue that the resultant PP account is

unacceptable because it allows for the conceptual distinction between physical and phenomenal which gives rise to the *hard problem of perceptual consciousness*. I diagnose this problem to be rooted in PP's rejection of *activity* and *knowledge-how*. Then, I explain why RSE's acceptance of these tenets allows it to provide a phenomenologically satisfactory explanation of conscious perception which does not allow for the 'hard problem'. Consequently, I conclude that RSE should be preferred over cognitivist accounts (like PP).

In chapter four I focus on the empirical phenomenon of *split-brain syndrome*. Split-brain surgery involves severing the corpus callosum and results in an anatomical disconnection between the left and right hemispheres of the brain. Split-brain patients are virtually indistinguishable from the general population in everyday life. However, in certain experimental situations, each of their hemispheres is capable of autonomously controlling goal-directed behaviour. As such, it appears as if the split-brain patient's conscious field 'splits' during these experiments.

In part one of chapter four I explain that cognitivist accounts of split-brain syndrome can be grouped into four different types, and that each of the four explanations fails for the same reason— they cannot account for the normal behaviour of split-brain patients outside of experimental contexts. Then, I argue that cognitivists are forced to accept one of these four accounts (in spite of their inability to explain the everyday behaviour of split-brain patients) because they reject *activity* and *knowledge-how*. These theorists do not necessarily believe that their account of split-brain syndrome provides the best explanation of the empirical data. Rather, they think their account should be preferred because it provides the best explanation of split-brain syndrome which is compatible with cognitivism.

In part two of chapter four I explain why acceptance of the two tenets allows for a more satisfactory explanation of split-brain syndrome, upon which consciousness is thought to split only during the experimental situation [*cf.* Hurley, 1998, ch. 5; 2003]. I provide my own RSE account of split-brain syndrome by arguing that split-brain patients learn a new sensorimotor skill— “use-external-factors-in-place-of-corpus-callosum”. As the name suggests, this strategy involves using external factors (that is, factors beyond the brain) as a functional surrogate for the now missing corpus callosum. Because RSE accepts the tenets of *activity* and *knowledge-how* it can allow for external factors to play a constitutive role in conscious perception, and so these external factors can serve to unify the subject’s conscious perceptual states. Consequently, RSE explains split-brain patients’ every day behaviour to be based on their having learnt new perceptual knowledge-how, and it explains their bizarre behaviour during experimental contexts to be based on the experimental controls blocking the exhibition of this knowledge-how. I motivate my account with reference to extant empirical evidence and suggest ways it could be empirically tested.

RSE can provide a satisfactory explanation of split-brain syndrome because it accepts the tenets of *activity* and *knowledge-how*. Cognitivist accounts cannot provide such a satisfactory explanation because they reject these tenets. Consequently, in chapter four I conclude that an RSE approach to conscious perception is to be preferred.

In chapter five I focus on the most common objection to theories of conscious perception predicated on *activity* and *knowledge-how*— that they fail to explain the role of the brain. Cognitivist approaches are generally thought to provide a good explanation of the brain’s role in conscious perception, and so are preferred on this basis. I argue that, when applied to RSE, this objection fails, and do so by focusing on the example of binocular

rivalry. I begin by outlining the predictive processing (PP) account of binocular rivalry, which is generally thought to provide a good explanation of the brain's role in conscious perception [Hohwy, Roepstorff, & Friston, 2008]. Then, I argue that this PP account of rivalry can be accommodated without accepting representation [Orlandi, 2015; *cf.* Orlandi, 2014]. Finally, I explain how this non-representational PP account could be subsumed within the RSE framework, and argue that a non-representational version of PP can be used to operationalise and explain the sub-personal aspects of RSE. I therefore conclude that RSE is capable of explaining the brain's role in conscious perception.

Although it may be granted that phenomena like binocular rivalry can be accounted for without requiring sub-personal representation, one could respond that PP explanations of other empirical phenomena make indispensable use of representation, and so realism about sub-personal representation is required after all. In chapter six I respond to this potential objection by arguing that, even if representation is *epistemologically* indispensable in *some* areas of PP research, this does not require that it *metaphysically* exists. I argue that representation can be understood as a theorist's fiction within PP, and apply a version of *make-believe* approaches to scientific models to PP in order to explain how representation could be playing an explanatory role in theorising despite not actually existing in PP systems. Consequently, I conclude that a non-representational version of PP shows promise of accounting for vast swathes of the empirical work carried out in this paradigm— either representation is rendered otiose, or it can be assigned a fictional status.

Finally, in chapter seven, I focus on 'non-veridical' experiences (dreams, hallucinations, and illusions). It is often argued that theories like RSE should be rejected because they cannot provide satisfactory explanations of these phenomena [Burge, 2005;

Block, 2005; Clark, 2012; Prinz, 2006; Seth, 2014]. By drawing on the work of William Fish [2009], I show that direct realist accounts like RSE can account for the existence of non-veridical experience. Although providing a full RSE account of such experiences (and pitting the resultant account against cognitivist rivals) goes beyond the scope of this thesis, I do show that the mere existence of such experiences does not provide any reason to be suspicious of RSE's ability to provide a general theory of consciousness.

In this thesis I develop and defend my own approach to conscious perception, which I label "radical sensorimotor enactivism" (RSE). RSE understands conscious perception to be an activity an organism performs which is predicated on its exercising the relevant knowledge-how. Perception occurs when sensorimotor knowledge-how is exercised by an organism, and conscious perception occurs when this sensorimotor knowledge-how is exercised attentively. RSE can provide an explanation of conscious perception which avoids the 'hard problem of consciousness' and it can provide a parsimonious explanation of split-brain syndrome, whereas cognitivist approaches cannot. Furthermore, by accepting a non-representational version of PP and applying it the sub-personal level of explanation, RSE provides an empirically tractable account of brain-based processing. I conclude that RSE provides a strong conceptual framework for the scientific study of conscious perception, and that it is worthy of further research and development in a number of areas of philosophy and cognitive science.

## Chapter One: A Rylean Cognitive Science

In his 1949 *The Concept Of Mind*, Gilbert Ryle mounted a systematic attack on a conception of mentality which he claims was invented by René Descartes. Descartes argued for the view known as *substance dualism*, according to which persons are composites of two different substances— *res extensa* and *res cogitans*. Descartes used the term “*res extensa*” to refer to material substances, and he argued that the defining characteristic of material substances was that they are necessarily spatio-temporally located. He used the term “*res cogitans*” to refer to the mind, and he defined the mind itself to be completely the opposite of material substances. For Descartes, the mind is not material (and so it has no spatio-temporal location) and its defining characteristic is thought. Ryle outlined a number of serious problems for Descartes’ concept of mind before arguing that, by adopting an anti-Cartesian view of mentality (developed by looking to how humans ordinarily speak about the mind), one arrives at a conceptual framework which satisfactorily explains mentality whilst avoiding the problems inherent in Descartes’ own view.

The two central tenets of Ryle’s account of mentality which will guide this thesis are as follows:

1. *Activity*— mental language acts as a place-holder for the capacities an organism possesses which allows them to engage in certain activities. Mentality is something that an organism *does*, it is not something which occurs passively within it.
2. *Knowledge-How*— an organism’s ability to engage in mental activity is predicated on its possession of the requisite *knowledge-how*. This ability should not be described in terms of *knowledge-that*.<sup>1</sup>

---

<sup>1</sup> In this thesis I assume that knowledge-how ≠ knowledge-that, because arguing for this point would require a thesis of its own. For arguments against the claim that knowledge-how ≠ knowledge-that the interested reader can consult [Stanley & Williamson, 2001; Stanley, 2011]. For arguments in favour of the claim they can consult [Anderson, 2014; Carter & Pritchard, 2015; Hutto, 2007; Hutto & Myin, 2013; Noë, 2012; 2015; Ryle, 1949/2000; 1946; Wittgenstein, 1953/2009].

In this thesis I adopt an approach to mentality which has been inspired by Ryle. I will argue that much of contemporary cognitive science (specifically, *cognitivist* cognitive science) is essentially Cartesian in nature. I will argue that the problems facing contemporary cognitive science can be avoided if one adopts a broadly Rylean approach to the mind, and do so by developing my own theory of conscious perception (radical sensorimotor enactivism) and explaining why it should be preferred over cognitivist theories.

### **§1: Cartesian Cognitive Science**

The *cognitivist* paradigm in cognitive science presents one with an essentially Cartesian view of the mind. Although cognitivists do not accept Descartes' dualism, they do explicitly reject the tenets of *activity* and *knowledge-how*. *Activity* is generally rejected by cognitivists because it is thought to provide an unsatisfactory explanation of mentality. A representative example of this line of thought is to be found in the work of Hilary Putnam, who focused on the example of "pain" [1975]. In order to provide an *active* definition of pain, one must define pain in terms of an organism's capacity to engage in pain-behaviour. For example, an *active* account of pain will define pain in terms of an organism's capacity to scream, move away from the painful stimulus, report that they are in pain, and so on and so forth. This account of pain is extremely revisionary. Pain behaviour is generally thought to occur *because* we feel pain, and behaviours such as screaming are thought to occur in *response* to the feeling of pain. Putnam presented an argument for this more common-sense view, and therefore concluded that pain is not constituted by pain-behaviour.

In his *Super-Spartan* thought experiment Putnam asks us to imagine a tribe of humans living in a Spartan culture. In this culture, any and all appearances of weakness are

to be avoided. As such, although people in this tribe do feel pain, they have been raised from birth to abhor the public exhibition of pain and avoid engaging in *any* kind of pain-behaviour. In short, Super-Spartans feel pain and yet they do not engage in pain behaviour. Therefore, this thought experiment shows that the feeling of pain is logically distinct from pain behaviour and yet *activity* appears to conflate the two. Consequently, *activity* is considered to be false.

The *knowledge-how* tenet is usually rejected by cognitivists on the basis of *poverty of the stimulus* (POS) arguments. Cognitivists argue that the information which organisms receive from the environment is too impoverished to explain successful cognition. Organisms can nonetheless cognise, and so it must follow that environmental information is supplemented with something internal to the organism. Cognitivists argue that the POS is overcome because environmental information is supplemented with information encoded in internal representations. The concept “representation” is therefore thought to provide a solution to the problem of explaining how successful cognition can occur in spite of impoverished environmental input. Representation cannot be explained in terms of knowledge-how, and so accepting representation requires that cognitivists reject the *knowledge-how* tenet.

Consider the example of vision. Contemporary cognitive science explains vision to be constituted by brain-processes, and the environment is thought to play the causal role of providing input for these brain-processes. Visual brain-processes begin at the eye, which is capable of picking up extremely sparse, two-dimensional information from the environment. When we perceive visually, however, we perceive three-dimensional visual scenes. The cognitivist is therefore faced with a visual POS argument— our brains receive



two-dimensional visual information but we perceive three-dimensional scenes. Cognitivists posit “representation” to solve this POS problem. The brain is thought to supplement its sparse retinal information with contentful internal representations, and these internal representations explain why we perceive three-dimensional visual scenes.

Contemporary cognitivist cognitive science therefore explicitly rejects Ryle’s two central tenets of mentality. *Activity* is rejected because it is argued to provide an unsatisfactory account of mentality, and *knowing-how* is rejected because it is incompatible with representation. In subsequent chapters I will highlight a number of serious problems facing the cognitivist research paradigm, and I will argue that these problems arise because cognitivists reject these two tenets.

## **§2: Rylean Cognitive Science**

In chapter two I will outline my own Ryle-inspired theory of conscious perception. In this section, I outline two extant approaches to cognitive science which are explicitly Rylean in character— *ecological psychology* and *situated robotics*. Each of these approaches is proposed as a distinctively empirical research programme which should be judged on the strength of its empirical results. Both approaches serve as exemplars of the approach to cognitive science which I intend my own theory of conscious perception to follow.

### **§2.1: Ecological Psychology**

*Ecological psychology* emphasises the active, direct, and temporally extended nature of cognition. Proponents of this research programme argue that an emphasis on these aspects

of cognition will lead one to the realisation that there is no poverty of the stimulus in most cognitive domains. If there is no poverty of the stimulus in a given cognitive domain, then representation should be rejected in that domain because it is posited as the solution to a non-existent problem [Anderson, 2014; Barrett, 2011; Chemero, 2009; Gibson, 1979].

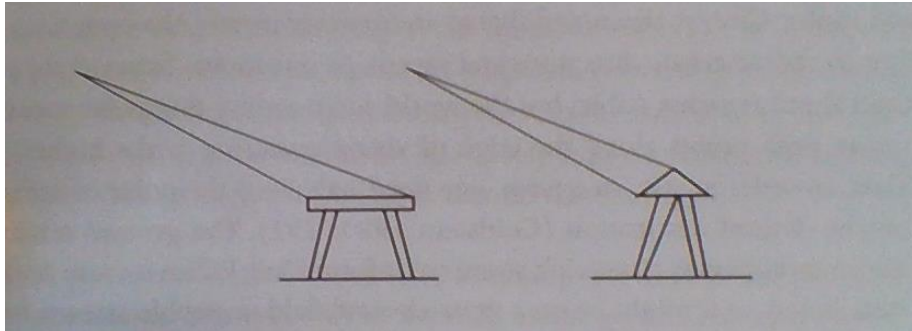
Michael Anderson makes this point vividly clear (in the case of vision):

When we attend to the fact that looking is an activity, that part of seeing is moving, we see that the true data of perception are extremely rich, multimodal, and perfectly capable of revealing the higher-order invariants in our environment and uniquely specifying the shape of the world. [Anderson, 2014, p. 169]

Ecological psychologists argue that perception involves a direct relation between the perceiving organism and its environment. Organisms perceive when they use environmental information to ‘pick-up’ or ‘uncover’ its *invariant properties*. Invariant properties are uncovered because organisms are *active* and perception is *temporally extended*. When an organism interacts with its environment, perceptually invariant properties of the environment will be uncovered, and in this manner organisms come to directly perceive objects in the world. I will use the example of vision to make this point clearer.

The following diagram presents an extremely simplified example of a poverty of visual stimulus scenario:

Figure One: *Sample Poverty of Visual Stimulus*

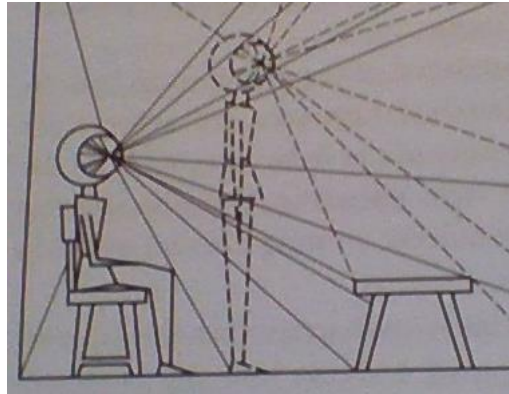


Shapiro, 2010, Fig. 2.3.

This diagram depicts how the visual input received by the organism's brain can be seen as compatible with two different environmental objects— a rectangular table and a pyramidal table. Cognitivists argue that because environmental information underdetermines the nature of the object being visually perceived, the brain must supplement this information with internal representations. For example, the brain may use prior representational knowledge that “tables are usually rectangular” and that “pyramidal tables are rare” and therefore hypothesise that a rectangular table is present. It is for this reason that a rectangular table will be visually perceived.

Ecological theorists argue that these kinds of explanation only appear to be required because their proponents fail to take into account the *active* and *temporally extended* nature of perception. Consider the following diagram:

Figure Two: Movement Disambiguates Visual Stimulus



Shapiro, 2010, Fig. 2.2. (taken from Gibson, 1979, Fig. 5.4)

This diagram depicts the same viewpoint of an organism in front of a rectangular table which was depicted in the previous diagram. However, it also depicts how the visual input received by the organism will change when it stands up and moves closer to the table. If the organism performs this action it will become obvious that the table is rectangular. The reason for this is that the visual input received over this *temporally extended interaction* with the environment is compatible only with rectangular tables. It is not compatible with pyramidal ones. Consequently, the organism's temporally extended perceptual interaction with the table serves to uncover the invariant properties of the rectangular table and uniquely specify it as such.

Ecological psychologists argue that, by taking into account the fact that organisms possess bodies and that they act in the world over temporally extended periods, it becomes obvious that there is no poverty of perceptual stimulus. If there is no poverty of the

stimulus, then there is no need to posit representation. Because they conceive of perception in this manner the ecological movement in psychology advances quite a specific research programme— study the perceptual invariants an organism is ‘attuned to’ or ‘uncovering’ when they perceive. Ecological psychologists argue that their theories should be advanced and defended on the strength of the empirical research programme they give rise to. One of their biggest successes in this area has been the study of how baseball fielders catch fly-balls.

In baseball, a “fly-ball” is a ball which has been batted high into the air. Professional baseball fielders are extremely good at catching fly-balls whereas non-fielders perform comparatively poorly. An obvious task for cognitive science would be to explain this difference — what is it that professional fielders do when they catch which makes them so good at it?

A natural cognitivist answer to this question will invoke the concept of “representation”. For example, the brain may be thought to use its representational capacities to calculate where the ball will land in the same manner that a physicist would. Once the most likely landing area has been calculated the organism will move to this place and catch the ball. The expertise of baseball fielders would then be explained as based on their brains having developed a more advanced and accurate representational heuristic or strategy than lay catchers.

It turns out, however, that the best empirical explanation of expert catching behaviour does not require representation. Rather than using representational capacities to catch fly-balls, it appears that fielders are instead tracking invariant properties in the environment. On one popular interpretation of the empirical data it appears that catchers

are using an *optical acceleration cancellation* strategy. By keeping the acceleration of the ball constant in their visual field the fielder will arrive at the correct place to catch the ball. If the ball appears to be accelerating in their visual field then this is a sign that they should speed up their own movements, because otherwise the ball will land on the ground before they are able to catch it. If it appears to be decelerating then they should slow down their movements, because otherwise they are going to arrive in the wrong place. Another popular interpretation of expert fielding is known as the *linear optical trajectory* strategy. This strategy takes expert catchers to move in such a way that the ball always appears visually to be moving in a straight line. If the ball visually appears to be moving in a straight line, then the catcher will arrive at the same place as the ball when it lands. Although it is still an open question which of these strategies are used by expert fielders, there is unanimous agreement that fielders are tracking invariants to catch fly-balls and that they are not using representational strategies [Shaffer & McBeath, 2002].

The example of research on catching fly-balls does not thereby refute cognitivism. But it does effectively demonstrate the approach to cognitive science suggested by ecological psychology— the study of cognition (in most domains) requires finding the invariants being tracked and understanding how tracking of these invariants leads to successful completion of the cognitive task. If organisms do in fact track environmental invariants in order to achieve a cognitive task, then there is no poverty of environmental stimulus and so representational explanations will not be required. Importantly, this ecological research programme is advanced as an empirical approach. The success of this research programme and the question of whether it should be preferred over cognitivism are both to be determined empirically.

It should be obvious that the ecological approach to mentality respects the tenets of *activity* and *knowledge-how*. Ecological theorists emphasise *activity* because cognition is conceptualised as an action performed by an organism. They emphasise *knowledge-how* because they reject representational explanations of cognition (in most cognitive domains). Respect of these two tenets has led to an empirically fruitful research programme. In this thesis I will argue that my own theory of conscious perception, which is also constrained by the tenets of *activity* and *knowledge-how*, will similarly lead to an empirically fruitful research programme which should be accepted on the basis of its empirical adequacy.

## **§2.2: Situated Robotics**

The study of robotics has traditionally been explicitly Cartesian (in the sense of Cartesian outlined in §1), and indeed the area of robotics known as “*good-old-fashioned artificial intelligence*” played a crucial role in founding the discipline of cognitivist cognitive science. Roboticists working within this representational tradition did have a lot of success creating robots which could successfully engage in intellectually demanding tasks, such as playing chess and performing mathematical calculations. However, they had minimal success developing robots which could perform practical tasks like walking or picking up objects. Frustrated with the lack of the success in this area, roboticists like Rodney Brooks and Randy Beer eschewed a cognitivist framework and instead built robots which did not use representation to complete practical actions. They met with remarkable success.

A representative example of this success is Rodney Brook’s robot “*Herbert*” [Brooks & Flynn, 1989]. Herbert was developed to navigate the MIT robotics lab, picking up empty

soft-drink cans and putting them in the bin. A traditional cognitivist approach to this task would involve constructing a representation of the external environment, using it to create a plan for action 'off-line', and then implementing this motor programme. Rather than designing Herbert in this cognitivist manner, Brooks provided him with a simple *subsumption* architecture which did not require or use representations. This simple architecture gave rise to an autonomous robot which was extremely effective at carrying out its job of collecting and disposing of cans.

Herbert's subsumption architecture was constituted by a number of simple modular layers, with each layer in charge of controlling a specific behaviour. One layer was controlled by a laser which would detect the presence or absence of a table in the vicinity. If a table was detected this laser initiated a forward movement which drove Herbert toward the table. Herbert would continue to move in the vicinity of the table until a laser controlled by a separate layer detected a soft-drink can. If a soft-drink can were to be detected by the laser, its detection would initiate an action which required movement toward the can. Herbert would move toward the can until the laser detected he was centred in front of it. At this point the laser would halt Herbert's behaviour and he stopped moving. Once Herbert had completely ceased moving a reaching mechanism was triggered. His arm would reach out toward the can and grasp it, with this reaching process itself determined by his subsumption architecture in the same manner as his arrival in front of the can.

We can see that Herbert's simple architecture gave rise to quite sophisticated behaviour and that it did so without requiring representation. Indeed, not only did Herbert's architecture not require representation, it seems that it was successful largely *because* it was not representational. This has been a general lesson from work in robotics. Robots with



non-representational architectures tend to be much better at practical activities, such as walking and grasping, than those with representational architectures [Barrett, 2011]. The conclusion reached by situated roboticists is that this success gives one license to infer that organisms do not use representation when they engage in practical cognitive activities.

The situated robotics research programme provides support for an approach to mentality which respects the tenets of *activity* and *knowledge-how*. *Activity* plays an essential role in these robots because it is only by acting and interacting with the environment that cognition can occur at all. Herbert's subsumption architecture, for example, requires activity for its success because if there is no activity we are left with a simple and inert modular architecture. However, when activity occurs and Herbert interacts with his environment, sophisticated autonomous behaviour emerges. Because these robots do not require representation their activity can be explained wholly in terms of *knowledge-how*. Once more, as was the case with ecological psychology, the situated robotics movement is advanced as an empirical research programme. Its proponents argue that it should be accepted on the basis of its empirical success, and not on the basis of a priori theoretical or philosophical assumptions.

### **§3: Conclusion**

In this thesis, I am going to provide an account of conscious perception which is inspired by Ryle's two tenets and which should be pursued in a manner similar to the research programmes of ecological psychology and situated robotics. I am going to argue that, by developing a theory of conscious perception which respects *activity* and *knowledge-how*, one arrives at a robust account of conscious perception which avoids many of the problems

faced by cognitivist accounts. Consequently, I will conclude that this Ryle-inspired theory of conscious perception should be preferred over cognitivist accounts as an empirical approach to the scientific study of conscious perception.

## Chapter Two: Radical Sensorimotor Enactivism

In this chapter I develop my own approach to conscious perception which has been inspired by, and developed in accordance with, the two tenets discussed in chapter one. I begin this chapter by outlining the *sensorimotor enactive* (SE) theory of conscious perception, upon which my own account is predicated. SE understands perception to be constituted by a direct relation between an organism and its environment. This direct relation is thought to be enabled by the organism's possession and exercise of *sensorimotor knowledge*, and conscious perception occurs when the organism attends to the exercise of this knowledge. Although SE respects *activity*, extant SE explanations of sensorimotor knowledge and attention require representation, and so extant versions of SE are incompatible with the tenet of *knowledge-how*. I will argue for a dispositional explanation of sensorimotor knowledge which dispenses with representation. Then, I will argue for an adverbial, non-representational theory of attention and explain how it can be applied to SE. Consequently, I will arrive at a thoroughly non-representational version of SE, and so at *radical*<sup>2</sup> sensorimotor enactivism. In subsequent chapters, I will apply radical sensorimotor enactivism to a number of issues in the contemporary cognitive science of conscious perception and argue that it provides a better explanation of conscious perception than its cognitivist competitors.

---

<sup>2</sup>Non-representational theories of cognition are often labelled "radical". This terminology was first introduced by Andy Clark [1997], who used the term to refer to theories of embodied cognition which dispense entirely with representation.

### **§1: What is Sensorimotor Enactivism?**

Sensorimotor enactivism received its first canonical explication and defence in O'Regan and Noë [2001]. In this paper, Kevin O'Regan and Alva Noë argued against cognitivist theories of perception and offered up a novel version of the ecological view in their stead. O'Regan and Noë's *sensorimotor enactive* (henceforth SE) theory was predicated on three central tenets:

1. Perception is relational
2. Perception is enabled by *sensorimotor knowledge*
3. Attention is necessary for *conscious* perception

SE is a relational theory of perception because perception is thought to be constituted by a direct relation between the perceiving organism and objects in its environment. SE takes organisms to come into direct perceptual contact with environmental objects when they possess *sensorimotor knowledge*. This is knowledge of the law-like relation between sensory stimulation and movement. For example, there is a law-like relation between visual stimulation and movement because visual stimulation changes in certain specific ways in tandem with our movements— as we move closer to visual objects they will loom in our visual field, as we move further away they will appear smaller, and so on. SE takes consciousness to be constituted by *access consciousness*.<sup>3</sup> A given state is access conscious if it can be used to govern thought, guide action, and (in linguistic organisms) inform linguistic report [Block, 1995]. SE takes attention to be necessary for conscious perception on the basis of empirical evidence from the change and inattention blindness paradigms. These paradigms appear to show that, unless an organism attends to a given aspect of their perceptual scene, they will not be conscious of it. For example, in change blindness

---

<sup>3</sup>The interested reader should consult [Cohen & Dennett, 2011; O'Regan, 2011; O'Regan & Noë, 2001], for defences of the view that access consciousness should be considered constitutive of consciousness.

experiments changes in a perceptual scene occur at the same time as a distractor stimulus. This distractor stimulus interrupts the normal functioning of the organism's attentional mechanisms, and the result is that large changes in a perceptual scene can go unnoticed. As such, the perceiver is not considered to have been conscious of the changes.

### **§1.1: SE and Activity**

SE is specifically advanced as an active view of perception, as the following quotes will attest:

Perception is not something that happens to us, or in us. It is something we do. [Noë, 2004, p. 1]

In the case of sponge squishing and Porsche driving, it seems clear that their accompanying feels are qualities of *what we do* when we squish the sponge and drive the Porsche. [O'Regan, 2011, p. 109, *italics in original*]

Consequently, SE clearly fulfils the tenet of *activity*. Although SE does straightforwardly fulfil this tenet, extant versions of the view do not obviously fulfil the *knowledge-how* constraint.

### **§1.2: SE and Knowledge-How**

Proponents of SE have always been keen to emphasise that sensorimotor knowledge is a *practical* ability possessed by organisms which is predicated on their possession of the requisite *knowledge-how*. However, the concept itself can appear to require (or at least invite) a representational reading. This point has been most forcefully argued for by Dan

Hutto [Hutto, 2005; *cf.* Hutto & Myin, 2013, pp. 23-32]. Hutto explains that the concept “sensorimotor knowledge” requires representation in order to play the explanatory role accorded to it by SE. If sensorimotor knowledge is understood non-representationally then, according to Hutto, it becomes trivial and loses its explanatory power. The biggest problem with extant explanations of sensorimotor knowledge, therefore, is that they either *require* representation or they are explanatorily vacuous.

A similar problem appears in extant SE treatments of attention. Although attention has always played a crucial role in SE, very little effort has been made by exponents of SE to explain how attention should be understood (both conceptually and empirically). Indeed, the only work carried out in this vicinity requires explaining attention in terms of *higher-order thought* theory [O’Regan & Noë, 2001; O’Regan, 2011]. Higher-order thought (HOT) theory is a theory of consciousness which takes a mental state to be conscious when the organism (or, more accurately, a part of the organism’s brain) possesses a higher-order thought about that state. If no higher-order thought is present, the mental state will be unconscious. HOT theory violates the knowledge-how constraint because it indispensably requires representation [Carruthers, 2016]. As such, extant versions of SE either explain attention by violating *knowledge-how* or they fail to explain attention at all.

## **§2: Radicalising Sensorimotor Enactivism**

In this section I will provide a non-representational account of both sensorimotor knowledge and attention. I will apply these non-representational accounts to SE later in the chapter, and therefore arrive at a non-representational version of the theory. Before doing

so, however, I must first define two key terms— the *sub-personal* and *personal* levels of description.

When I use the term “*sub-personal level of description*” in this thesis I take myself to be referring to biological processing which occurs within the brain.<sup>4</sup> For example, if I provide a sub-personal level description of my visit to the coffee shop, I will describe the processes which occurred in my brain during the coffee shop visit. I will refer to neurons, to white and grey matter, to the secretion of neurochemicals, to the propagation of electrical voltages, and so on [*cf.* Dennett, 1969/2010; McDowell, 1994].

When I use the term “*personal level of description*” in this thesis I take myself to be referring to organismal actions. In the case of human beings, the personal level of description is concerned with the everyday activities which human beings engage in. For example, the activities of drinking coffee, walking through the park, and starting an exercise programme all belong to the personal level of description. These are activities which are properly ascribed to the agent itself, and not parts of that agent [Bennett & Hacker, 2003; Ryle, 1949/2000; Wittgenstein, 1953/2009].

### ***§2.1: Non-Representational Sensorimotor Knowledge***

In order to definitively avoid a representational understanding of sensorimotor knowledge, I propose that it should be understood wholly in terms of dispositions. At the sub-personal

---

<sup>4</sup>In providing this definition, I am not stipulating or requiring that sub-personal processing can *only* occur within the biological brain. Indeed, later in the thesis I will argue that sometimes sub-personal perceptual processing constitutively includes extra-neural factors [ch. 3, pt. 2, this thesis; *cf.* Clark & Chalmers, 1998]. However, when I use the term “sub-personal” in this thesis I should be taken to be referring to the brain unless I explicitly note otherwise.

level of description, I propose that sensorimotor knowledge be understood in terms of dispositional properties of the brain, whilst at the personal level I suggest that it be understood in terms of dispositional properties of the organism.

Psychological behaviourists explained mentality solely in terms of sensory input and behavioural output, arguing that mentality is constituted by relations between certain sensory inputs and certain other behavioural outputs (or a series of different behavioural outputs, depending on the organism's phylogenetic and ontogenetic history). The brain itself was therefore understood as a dispositional system, which provided causal mediation between sensory inputs and their resultant behavioural outputs. By applying a psychological behaviourist approach to sensorimotor knowledge at the sub-personal level of explanation, one could arrive at a thoroughly non-representational explanation of it [*cf.* Block, 2001]. Explaining sensorimotor knowledge in this manner would require taking it to describe nothing more than a series of causal relations between certain sensory inputs and certain other motor outputs. The brain, therefore, would be considered to play the role of a causal mediator, and causal mediators should not be described in terms of representation [*cf.* Hutto & Myin, 2013, ch. 4; Ramsey, 2009, ch. 4].<sup>5</sup>

At the personal level of explanation, I propose that Gilbert Ryle's theory of knowledge-how be applied to sensorimotor knowledge. Ryle argued that an organism's ability to perform a given action is predicated upon its possession of the relevant knowledge-how. He understood knowledge-how itself to be constituted by the dispositional properties of an organism. As such, if an organism knows-how to perform a given activity, it

---

<sup>5</sup> In chapter five of this thesis I discuss a possible empirical operationalisation of this idea (via predictive processing). Once this empirical operationalisation has been provided, I will have arrived at an account of sensorimotor knowledge which is both non-representational and genuinely explanatory.



will exhibit this know-how in its actual behaviour and counter-factual behavioural tendencies [Ryle, 1949/2000, ch. 2]. For example, an organism that knows-how to throw stones will exhibit stone-throwing behaviour in appropriate scenarios (scenarios in which stones are available, the organism deems it pertinent to throw stones, and so on). By applying Ryle's account of knowledge-how to personal level sensorimotor knowledge, we will therefore ascribe sensorimotor knowledge when an organism engages in, or is disposed to engage in, perceptual behaviour. The organism will, for example, exhibit mating behaviour when it *sees* another member of its species of the requisite sex, when it *smells* the presence of such a member, and so on and so forth. Consequently, we can take an organism to possess sensorimotor knowledge if it is disposed to behave in a manner which is consistent with its understanding the law-like relation between sensation and movement. Because knowledge-how is explained in terms of non-representational dispositions, this explanation of sensorimotor knowledge is compatible with the *knowledge-how* tenet.

In short, sensorimotor knowledge can be understood solely in terms of non-representational dispositions at both the sub-personal and personal levels of explanation. Given that I will advocate an approach to conscious perception which respects *knowledge-how*, I therefore conclude that sensorimotor knowledge both: should be understood in non-representational terms; and, that it can be so understood.

## **§2.2: Non-Representational Attention**

In this sub-section I am going to suggest that “attention” on the SE framework is best understood as an adverbial phenomenon. I will first outline Chris Mole's metaphysical

argument for adverbialism about attention. Then, I will explain how Mole’s adverbial theory of attention can be modified in order to cohere with the *knowledge-how* constraint. The resultant non-representational theory of attention I arrive at will be applied to SE in section three of this chapter.

### **§2.2.1: Distinguishing “Process” and “Adverb”**

In his 2011 *Attention is Cognitive Unison* Chris Mole argues for the thesis that attention is an adverbial phenomenon. Mole’s argument is premised on there being a metaphysical distinction between “process” and “adverb”:

A taxonomy is a taxonomy on the basis of process if the taxonomy classifies events on the basis of having or gaining of a property *by an object*. A taxonomy is a taxonomy on the basis of manner if the taxonomy classifies events on the basis of the having or gaining of a property *by an event*. [Mole, 2011, p. 29, *italics in original*]

He argues that, in order to determine the metaphysical category of some *x*, we must consider the following two questions [2011, ch. 2]:

1. What is *x*?
2. What does it mean for *x* to occur?

Consider the examples of “combustion” (a process) and “hastily” (an adverb). In order to determine the metaphysical category of combustion it is natural to answer question one first. Combustion occurs when an object gains the property of burning because when an object combusts a chemical reaction between oxygen and fuel occurs which results in

burning. We can therefore classify combustion as a process which occurs to an object and causes it to gain the property of burning. Having arrived at an answer to question one, the answer to question two becomes obvious. For combustion to occur we require that the chemical process of burning occur. As such, Mole labels combustion a *process-first* phenomenon— in order to determine what combustion is we have to first understand what the *process* of combustion is.

Adverbs are words that describe the manner in which activities and events occur. Consider the adverb “hastily”. This adverb can be used to describe many different types of event— the publication of a newspaper can be performed hastily, a person’s walk to the train station can be performed hastily, and the actions of two particular Hobbits in Middle Earth can be performed hastily (from the perspective of a disapproving Ent). Each of these events involve entirely different processes and it would be difficult, if not impossible, to find a process which all of these hastily executed events have in common. Consequently, one cannot determine whether *x* was hastily executed solely by focusing on the process of *x*. Each of these events are similar because they are carried out in a similar manner (hastily), and not because they involve the same processes.

Consider the example of newspaper publication. In order to determine whether a newspaper was published hastily one should answer question two first. The reason for this is that, although there is a difference between a hastily published newspaper and a leisurely published one, this difference is not to be found by considering the process of publication itself. Both hastily and leisurely published newspapers follow exactly the same processes— reporters write articles, editors edit them, the publisher presses them onto paper, and so

on. As such, if we are to determine whether a publication was performed hastily, we need to understand the manner in which the process of newspaper publication was carried out.

In summary— Mole argues that there is a difference between certain objects and events, and that this difference can be captured by the metaphysical categories of “process” and “adverb”. Process-first phenomena can be grouped into the same set because they all involve an object gaining a property. Adverbial phenomena can be grouped into the same set because they all involve an event gaining a property. In order to determine the metaphysical category of a given  $x$ , one must determine whether it makes more sense to ask first of  $x$  what processes it involves, or instead to ask first of  $x$  the manner in which its processes are carried out.<sup>6</sup>

### ***§2.2.2: An Adverbial Theory of Attention***

I am now going to outline Mole’s application of his metaphysical taxonomy to the case of attention. According to Mole, in order to determine the metaphysical category of attention we must ask the following two questions:

1. What is attention?
2. What is it for something to be done attentively?

[Mole, 2011, p. 24]

---

<sup>6</sup> There may be cases wherein it is unclear whether a given  $x$  is best described as a process or an adverb [Watzl, 2011]. Although this is true, provided that there are clear-cut cases of process and adverb (such as combustion and hastily) I do not need to concern myself with this problem. My aim here is only to outline Mole’s metaphysical taxonomy— it is not necessary for me to fit each and every concept into this taxonomy.

Philosophers and psychologists have traditionally answered question one first.

Consequently, they tend to believe that attention is a process-first phenomenon. The consensus view that attention is a process has not, however, led to a consensus opinion on what attention is. There are so many different candidates for explaining the process of attention that most psychologists do not believe there is a single set of processes which are necessary and sufficient for attention. In order to show that attention is not identical to a given process, one must provide examples of the presence of the process in the absence of attention.

Consider the feature-binding theory of attention. Different aspects of perceived objects are generally thought to be processed in different parts of the brain. For example, colour is thought to be processed in V4, motion in V5, and so on. However, we perceive these various attributes as 'bound' onto the one object. Rather than experiencing bare particulars of colour and motion, we instead perceive a swirling green leaf. Anne Treisman has argued that attention is the mechanism responsible for binding the spatially and temporally distributed aspects of this neural processing together, such that we perceive objects instead of bare particulars. In addition to explaining 'normal' object perception, Treisman also takes her theory of attention to explain pathologies of object perception (see [Treisman, 1998] for details). This *feature-binding theory of attention* is a process-first theory because it understands attention to be "wholly constituted by feature-binding processes" [Mole, 2011, p. 37].

Mole uses the empirical phenomenon of *unilateral neglect* to argue that attention is not constituted by the process of feature-binding. Unilateral neglect is a neuropsychological disorder in which patients have trouble attending to one side of their visual field (usually their left side). Patients will, for example, only eat from one side of their plate, shave only

one side of their face, draw on only one half of a page, and so on and so forth. Although these patients cannot attend to one half of their visual field, feature-binding does occur in this unattended field. Consequently, attention cannot be identical to the process of feature-binding because in unilateral neglect feature-binding occurs but attention is not present. Attention therefore is not “wholly constituted by feature-binding processes”. Although Mole focuses only on the example of feature-binding, he does think that exactly this type of argument can be applied to most (if not all) process-first theories of attention.<sup>7</sup>

Having argued against process-first theories of attention, Mole offers in their stead his own positive proposal about how attention should be classified. In order to determine the metaphysical category of attention, Mole claims that we are better asking question (2) first. Therefore, he concludes that attention should be understood as an adverb. He then offers his own theory of attention, *Cognitive Unison* theory, which proposes that attention occurs when an organism uses its cognitive resources in unison to perform a task.<sup>8</sup>

Mole argues that, if an organism is to count as performing a task, the following three conditions must be met [Mole, 2011, pp. 52-55]:

1. The task must include the organism.
2. The organism must know-how to perform the task.
3. The organism must be putting their know-how to use.

---

<sup>7</sup> Although I will not provide additional examples in this thesis, I think Mole is on strong dialectical ground here. There is a complete lack of agreement within the field on which processes are constitutive of attention, and this lack of agreement is present largely because analogues of the argument Mole applies to feature-binding theories have been found for almost all theories of attention. For any given process proposed as constitutive of attention, examples have been found of the presence of this process in the absence of attention.

<sup>8</sup>With tasks themselves defined in terms of the specific way of life of the attending organism. A lioness, for example, can engage in the task of ‘hunting’ because its way of life involves such a task.

Consider a lioness which is attending to the task of hunting. In order to count as ‘attentively hunting’, the lioness herself must be engaged in hunting. Consequently, the task of hunting must be applicable to the individual lioness herself, as opposed to a chemical or sociological level of description. The lioness can only count as engaged in the task of hunting if she possesses the ability to hunt. Consequently, the ‘attentively hunting’ lioness must know-how to hunt. Finally, the lioness only counts as engaged in the task of ‘attentively hunting’ if she is currently putting her hunting know-how to use. If the lioness is asleep, for example, then she is not putting her know-how to use and so cannot count as engaged in the task of hunting.

In summary— Mole argues that attention occurs when an organism uses its cognitive resources in unison to attend to task *x*. Organisms will only count as ‘attentively *x*-ing’ if it is the organism itself which performs *x*, the organism knows-how to *x*, and the organism is currently engaged in *x*-ing. If these three conditions are met, then the organism can be ascribed the adverbial property of ‘attentively *x*-ing’.

### ***§2.3: ‘Radicalising’ Cognitive Unison***

Mole’s adverbial theory of attention is compatible with, and advanced in the spirit of, the tenet of *activity*. If attention is an adverb, then it is necessarily performative because it describes the manner in which an organism engages in a task. Although Cognitive Unison theory is compatible with *activity*, it is not compatible with the *knowledge-how* constraint. The reason for this incompatibility is that Mole takes his theory to indispensably require representation at both the personal and sub-personal levels of explanation.

Given that it is the organism itself which must attend to  $x$  in order to count as 'attentively  $x$ -ing', Cognitive Unison theory is applicable to the personal level of explanation. Therefore, attention is considered to be a type of personal level cognition. Mole contends that personal level cognition is itself to be understood in terms of representation:

A cognitive process, in this sense, is a process that operates on representations that encode their contents *for the agent of the task*: In order for a process to count as cognitive, there must be an agent-level contentful state whose content is directly determined, at least in part, by the content of the representations on which that process operates. [Mole, 2011, pp. 57-58, *italics in original*]

At the level of sub-personal empirical implementation, Mole argues that his theory of attention is compatible with *biased-competition* models of attention:

If the cognitive unison theory gives us the correct account of *what* attention *is*, then the biased-competition model may give us the correct account of *how* many instances of this attention-realizing unison *come about*, and of how they get maintained. [Mole, 2011, p. 133, *italics in original*]

Competition models of cognition take different areas of the brain to be engaged in constant competition with one another, with their end goal thought to be that of controlling neural processing. Biased-competition theory is a theory of attention which explains neural competition in terms of brain-based representations vying for control of the brain [Desimone & Duncan, 1995]. Consequently, biased-competition is a representational theory of attention, and so Cognitive Unison appears to require representation at the sub-personal level of empirical implementation.



Mole argues that Cognitive Unison requires a representational gloss at both the personal level of explanation and at the sub-personal level of empirical implementation. Consequently, his theory is incompatible with the *knowledge-how* constraint. In order to provide a non-representational version of Cognitive Unison theory, and so make it compatible with *knowledge-how*, I must therefore: explain personal level cognition in non-representational terms; and, explain how Cognitive Unison could be empirically implemented without requiring representation.

I am not going to provide a non-representational theory of personal level cognition in this thesis, because doing so would require (at least) a separate thesis of its own. Instead, I am going to assume that a non-representational theory of personal level cognition *could* be given (see, for example, [Anderson, 2014; Barrett, 2011; Chemero, 2009; Hutto & Myin, 2013; Ramsey, 2009]).<sup>9</sup> Given that there are plenty of non-representational theories of personal level cognition present in the literature, and given that Cognitive Unison is itself a type of personal level cognition, it therefore follows that Cognitive Unison *can* be accepted without requiring personal level representation. Therefore, provided one accepts a non-representational theory of cognition, one can arrive at a non-representational version of Cognitive Unison theory.

Where the empirical implementation of Cognitive Unison is concerned, although biased-competition theories are representational theories, there are non-representational

---

<sup>9</sup> Note, I am not assuming the truth of such non-representational theories of cognition. Rather, I am claiming that non-representational theories of cognition *can* be provided. Even the staunchest representational theorist has to agree on this point. Otherwise, they risk making representational theories of cognition correct a priori, and so weakening their empirical credentials [Ramsey, 2015].

alternatives which could be applied to Cognitive Unison. One such approach is to be found in the work of Michael Anderson [Anderson, 2014, esp. chs. 5 & 6; Anderson, 2015].

Anderson is an ecological psychologist who advances an account of mentality which respects the *activity* and *knowledge-how* constraints. In his 2014 *After Phrenology*, he argues for an extremely anti-modular, *neural-reuse* theory of cognition, upon which different neural regions are constantly altering their connections with one another in response to various task-demands. Anderson argues that brains are constantly engaged in the formation and dissolution of what he labels “*transiently-assembled-local-neural-subsystems*” (or TALoNS, for short). Although TALoNS can be studied via the traditional methods of systems neuroscience (for example [Sporns, 2010]), Anderson argues that we can come to a better understanding of TALoNS if we adopt the *affordance competition hypothesis*.

“*Affordance*” is a theoretical concept which states that organisms perceive their environment solely or essentially in terms of the activities it affords [Gibson, 1979]. Chairs, for example, afford sitting-behaviour to humans. Consequently, according to affordance theory, humans will perceive chairs as ‘sit-upon-able’. The affordance competition hypothesis proposes that organisms are constantly engaged in processing multiple environmental affordances at any given time— affordances compete with one another within the brain, with the winning affordance being the one that ends up controlling behaviour [Cisek, 2007].

Anderson argues that the brain’s neural dynamics should be studied in terms of affordance competition. By conceptualising neural dynamics in this manner, he contends that we can come to an understanding of why TALoNS are constantly being formed and re-

formed ‘on the fly’.<sup>10</sup> TALoNS are formed ‘on the fly’ because different affordances are competing with one another within the brain, with the winning affordance being the one which determines the nature of the TALoNS being deployed. He labels his theory “*biased-affordance-competition*”, and is quite explicit that it should not be understood in terms of representation. Because Anderson provides a non-representational version of biased-competition theory, we can apply his framework to Cognitive Unison theory and therefore arrive at a non-representational empirical implementation of it.

I began this sub-section by noting that although Cognitive Unison theory fulfils the *activity* constraint, it does not fulfil the *knowledge-how* constraint because it appears to require representation. Mole takes Cognitive Unison to require representation because: it is a type of personal level cognition and Mole thinks personal level cognition must be explained in terms of representation; and, it is thought to be implemented via biased-competition theory, and biased-competition theory requires representation. I explained that there is no reason to think one could not adopt a non-representational explanation of personal level cognition, and I used Michael Anderson’s work to show how one could provide a non-representational version of biased-competition theory. Consequently, I have shown that Cognitive Unison theory can be endorsed without accepting representation. We have, therefore, arrived at a ‘radical’ version of Cognitive Unison.

---

<sup>10</sup> I derive my application of this phrase to Anderson’s framework from Andy Clark [2016].

### **§3: ‘Radical’ Sensorimotor Enactivism**

In section one I explained that extant SE explanations of sensorimotor knowledge and attention either conflict with the *knowledge-how* constraint or the posits themselves are left entirely unexplained. In section two I provided a positive explanation of both of these posits which did not require representation. In this section I will apply these non-representational modifications of sensorimotor knowledge and attention to SE, and so arrive at a non-representational (hence *radical*) version of the theory. I call this theory “*radical sensorimotor enactivism*” (RSE).

RSE is a relational theory of perception which takes perception to be constituted by a direct relation between the perceiving organism and its environment. This direct perceptual relation occurs when an organism both possesses and exercises sensorimotor knowledge. Sensorimotor knowledge itself concerns knowledge of the law-like relation between sensory stimulation and movement, and is to be understood solely in terms of dispositions (at both the sub-personal and personal levels of explanation). Consequently, it should not be explained in terms of representation. Finally, RSE takes attention to be necessary for consciousness. Therefore, an organism will be conscious of its perceptual relation to the environment only when it *attends* to the occurrent exercise of sensorimotor knowledge. I propose that the non-representational version of Cognitive Unison (outlined in §2) should be used to explain what, exactly, it means for an organism to attend to the occurrent exercise of sensorimotor knowledge.

In order to count as attending to the task of perceiving, or as ‘attentively perceiving’, the organism in question must fulfil Mole’s three task criteria:

1. The agent must perform the task— RSE fulfils this criterion because this theory requires that it is the organism itself which directly perceives its environment.
2. The agent must know-how to perform the task— RSE explains perception as enabled by the possession of sensorimotor knowledge. If an agent possesses sensorimotor knowledge, then they will count as knowing-how to perceive.
3. The agent must be putting their know-how to use— RSE takes perception to occur when an organism is engaged in the occurrent exercise of sensorimotor knowledge. If the agent were asleep, for example, they would not count as occurrently exercising sensorimotor knowledge.

We can see that RSE straightforwardly fulfils Mole's three task criteria. Therefore, RSE can allow for an organism to be considered to attend to the task of perception and so allows for 'attentive perception'.

In order to better appreciate the view, it helps to consider an example. Imagine the following scenario— two people who take the exact same route whilst walking through a park, and who carry out the same perceptual processes. One of the walkers, who I will label the "*mindful-walker*", is walking through the park whilst practising mindfulness meditation. This being the case, they are consciously aware of their various perceptual relations to the environment as they make their way through the park. The mindful-walker can see the bright red of the rose in front of them, feel the cool breeze on their neck, and hear the gentling rustling of leaves behind them. The other walker, although taking the exact same route through the park, is not consciously aware of their perceptual relation to the environment. This *worrying-walker* is currently in the midst of a financial crisis and they are trying to work out how best to extricate themselves from their distressing predicament.

Although the worrying-walker is perceptually related to the red rose, the cool breeze, and so on, they are not conscious of this perceptual relation. Because both walkers take the same routes and engage in the same perceptual processes, each walker can be said to perceive. This explains why neither walker trips over branches, why both divert their gaze when they (accidentally) look straight at the sun, and so on and so forth. However, although both walkers are perceptually related to their environment, only one walker is conscious of their perceptual relation— the mindful-walker.

RSE explains the difference between each walker in terms of attention: the mindful-walker is engaged in ‘attentive perception’ whereas the worrying-walker is not. Whilst the former is using their cognitive resources in unison to attend to the task of perception, the latter is using their cognitive resources in unison to attend to their financial problems. Thus, the difference between the two walkers, and the reason why one is conscious of their perceptual relation whilst the other is not, is an adverbial one. It is not concerned solely with the processes of perception performed by both walkers. Rather, the difference between the two walkers concerns the manner in which their perceptual processing is carried out. Although both walkers are engaged in the same perceptual processing, only one of these walkers (the mindful-walker) carries out this perceptual processing ‘attentively’.

#### **§4: Conclusion**

In this chapter I developed my own radical sensorimotor enactive approach to conscious perception. In section one I outlined the theory of conscious perception upon which my own account is founded— sensorimotor enactivism. I noted that although sensorimotor enactivism fulfils the *activity* constraint, it does not fulfil the *knowledge-how* constraint

because: extant sensorimotor enactive theories have not explained how sensorimotor knowledge could be made compatible with *knowledge-how*; and, extant sensorimotor enactive theories have either explained attention in a manner which violates *knowledge-how* or not explained attention at all. In section two, I explained how sensorimotor knowledge could be understood in non-representational terms and outlined a non-representational version of Chris Mole's Cognitive Unison theory of attention. In section three I applied these non-representational accounts of sensorimotor knowledge and attention to sensorimotor enactivism and so arrived at a *radical* version of the view. In the rest of the thesis I will argue that radical sensorimotor enactivism is to be preferred over cognitivist alternatives because it provides a better explanation of conscious perception at both the personal and sub-personal levels of explanation.

## Chapter Three: Cognitivism, Radical Sensorimotor Enactivism, and the ‘Hard Problem of Perceptual Consciousness’

In this chapter I argue that cognitivism is problematic because it enables the ‘hard problem of perceptual consciousness’. I explain that cognitivism faces this problem because it explicitly rejects *activity* and *knowledge-how*. Then, I argue that RSE avoids this ‘hard problem’, and I explain that it does so because it accepts *activity* and *knowledge-how*. Consequently, I conclude that we have strong reason to prefer RSE as a theory of conscious perception over cognitivist competitors.

The chapter is structured as follows— in section one I distinguish between the sub-personal, personal, and conscious levels of description. Then, I explain how cognitivist accounts propose to explain the inter-relation between these levels of description by focusing on the example of predictive processing. In section two, I argue that the cognitivist explanation of the inter-relation between these levels of description is unsatisfactory because it gives rise to the ‘hard problem of perceptual consciousness’. I explain that this problem arises because cognitivists reject *activity* and *knowledge-how*. Then, in section three, I provide an RSE explanation of the inter-relation between the three levels of description. I argue that RSE can provide a phenomenologically plausible explanation of the inter-relation between these levels which does not give rise to the ‘hard problem of perceptual consciousness’, and argue that it can do so because it accepts the tenets of *activity* and *knowledge-how*. RSE can explain conscious perception whilst avoiding the ‘hard problem of perceptual consciousness’. Cognitivist accounts, however, cannot. Therefore, I



conclude that this gives us strong reason to prefer RSE over cognitivist accounts of conscious perception.

### **§1: The Sub-Personal, Personal, and Conscious Levels of Explanation**

We encountered the sub-personal and personal levels of explanation in chapter two. I defined the sub-personal level of explanation as concerned with biological and chemical processing which occurs within brains, whilst the personal level of description was defined as concerning embodied agents that interact in a world of ‘medium-sized dry goods’.<sup>11</sup> I will now introduce a third level of description— the *conscious* level of description. When I refer to the conscious level of description, I take myself to be referring to ‘what it is like’ to experience the world. For example, there is ‘something it is like’ to see a beautiful painting, to hear a delightful melody, or to feel a luxurious garment.

There is quite an obvious disconnect between these levels of description, and it is the job of cognitive science to explain how they inter-relate. I am going to argue that cognitivist theories cannot explain this inter-relation without facing the ‘hard problem of perceptual consciousness’ by focusing on the example of predictive processing. I will then argue that theories of conscious perception which respect the tenets of *activity* and *knowledge-how* can explain this inter-relation without having to face the ‘hard problem’, and will do so by focusing on the example of RSE.

---

<sup>11</sup> This phrase was coined by the philosopher John Langshaw Austin.

### **§1.1: What is Predictive Processing?<sup>12</sup>**

*Predictive processing* (henceforth, PP) is a cognitivist theory of mentality which explains (conscious) perception as follows— the brain is an inference machine which uses Bayesian probability theory (or approximations thereof) to determine the most probable causes of the sensory stimulation it receives. It infers the nature of events external to the brain, and its ‘best inference’ will determine what it is that an organism (consciously) perceives.

PP is taken to provide a novel solution to *the problem of perception*, which is a poverty of stimulus argument specific to perception. Gilbert Ryle provides a beautiful statement of this problem:

There is immured in a windowless cell a prisoner, who has lived there in solitary confinement since birth. All that comes to him from the outside world is flickers of light thrown upon his cell-walls and tappings heard through the stones; yet from these observed flashes and tappings he becomes, or seems to become, apprised of unobserved football-matches, flower-gardens and eclipses of the sun. How then does he learn the ciphers in which his signals are arranged, or even find out that there are such things as ciphers? How can he interpret the messages which he somehow deciphers, given that the vocabularies of those messages are the vocabularies of football and astronomy and not those of flickers and tappings? [Ryle, 1949/2000, p. 212]

In order to solve the problem of perception one must explain how the brain (the “prisoner in a windowless cell”) could come to possess knowledge about the nature of external events

---

<sup>12</sup> This section is largely based on Jakob Hohwy’s excellent and extremely accessible introduction to predictive processing [Hohwy, 2013, Part 1].

(such as “football matches” and “flower gardens”) despite the fact that the stimulation received by the brain (“the flickers of light thrown upon its cell-walls”) under-determines the nature of these external events. The PP solution to this problem is to argue that the brain uses Bayes’ theorem (or approximations thereof) to infer the nature of events external to the skull.

Bayes’ theorem is predicated on three key mathematical concepts:

1. Likelihood: the likelihood of a given event, given the current information possessed by the model.
2. Prior: the prior probability of a given event when the information currently possessed by the model is ignored.
3. Posterior Probability: the overall likelihood that a given event is occurring. This is calculating by multiplying (1) and (2).

Consider as an example my present perceptual situation. I am currently sitting in a café in Brighton which is called “*The Laughing Dog*”. This café has a resident Golden Retriever (presumably, this dog is the *Laughing Dog*) which is currently ambling through the café, and so I am receiving visual stimulation which is consistent with the presence of a dog. Let us imagine that my Bayesian brain is attempting to infer the external causes of my visual stimulation, and that it has settled upon the following two mutually exclusive hypotheses:<sup>13</sup>

*a*: I am visually perceiving a Golden Retriever

*b*: I am visually perceiving a Golden Retriever hologram

---

<sup>13</sup> In practice things will be much more complicated than this. I am making a deliberate simplification here because my aim is only to provide an example of how PP proposes to solve the problem of perception.

Given that I am sitting in a café which has a resident Golden Retriever and that I am currently receiving visual stimulation consistent with the presence of a Golden Retriever, there is a high likelihood that I am being visually confronted with a Golden Retriever. Furthermore, the prior probability that a Golden Retriever will be walking around the *Laughing Dog* café is extremely high. The sum of the likelihood and the prior of hypothesis *a* will be extremely high, and so hypothesis *a* is going to have a high posterior. It is therefore highly likely that my visual stimulation is caused by the presence of a Golden Retriever.

My current visual stimulation is also extremely consistent with my being presented with a hologram of a Golden Retriever, and so the likelihood of hypothesis *b* will be quite high. However, the prior probability that I would be confronted with a holographic projection of a Golden Retriever is extremely low. In general, I do not frequent environments in which holographic projections are common, and most cafés do not operate holographic projections. Because of this low prior, the posterior probability of hypothesis *b* will be very low. It is extremely unlikely that I am perceiving a Golden Retriever hologram.

Hypothesis *a* has a much higher posterior than hypothesis *b*. Consequently, the best explanation of my current sensory stimulation is that I am visually perceiving a Golden Retriever. This hypothesis will result in my current perceptual experience, and it will also determine my expectations for further perceptual experiences. For example, I will expect to feel fur when I reach out to pet the dog, I will expect that my hand will not pass through the dog when I touch it, and so on. If hypothesis *b* were determining my perceptual experience, then I would not expect to feel fur when I pet the holographic dog, I would expect my hand to pass through it when I attempt to touch it, and so on. The expectations derived from the

currently dominant hypothesis will, therefore, subsequently guide and determine future action and behaviour.<sup>14</sup>

PP understands the brain to be constantly engaged in the creation and running of these sorts of probabilistic models, which create hypotheses about the most likely cause of the sensory stimulation being received by the brain. The brain's primary aim, on this theory, is to provide the best possible hypothesis about the causes of its current sensory stimulation. The brain is thought to achieve this aim by continuously attempting to minimize *prediction error*. Prediction error-signals signal that a given prediction is false— the sensory signal received by the brain was not the sensory signal which it predicted would be received. The brain is considered to be constantly updating its models in response to error-signals. It uses two different methods to minimize error: *perceptual inference* and *active inference*.<sup>15</sup>

*Perceptual inference* involves the brain revising its hypotheses to fit the sensory stimulation being received. For example, if the brain's hypothesis that "a dog is present" generates error-signals, the brain can revise its hypothesis to "there is no dog present". By revising its hypothesis in this manner the brain will be able to reduce the prediction error generated. *Active inference* occurs when the brain uses movement to change the sensory stimulation being received. For example, if the brain's hypothesis "a dog is present" generates error-signals, then the brain can use movement to reduce prediction error. It can cause the organism to continue moving until it comes into sensory contact with a dog, at

---

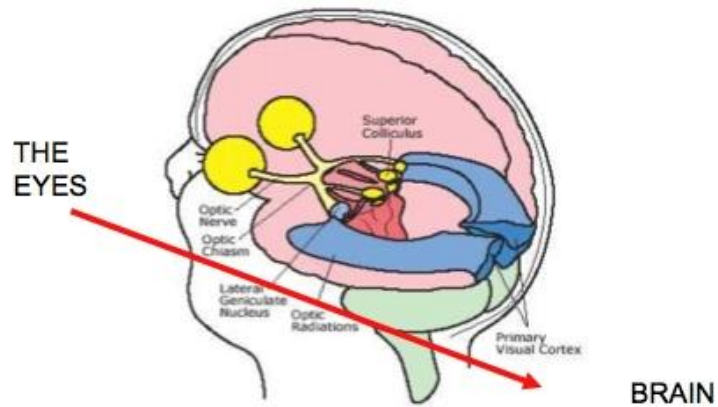
<sup>14</sup>It is for this reason that PP models are taken by proponents of the theory to be essentially *action-oriented* [Clark, 2013; 2016; Friston, 2012; Hohwy, 2013; Seth, 2014].

<sup>15</sup> The labels "perceptual inference" and "active inference" are used by Jakob Hohwy [2013] to distinguish between the two ways of minimising error. Theorists like Karl Friston and Andy Clark, however, use the term "active inference" to refer to the minimisation of error, and would label the two ways of minimising error "perception" and "action" respectively. Because nothing of importance (in the present context) hangs on this terminological difference, I will make use of Hohwy's labels in the main text.

which point the hypothesis “a dog is present” will be fulfilled. In short, perceptual inference involves changing the brain’s model to fit its sensory stimulation whereas active inference involves changing the brain’s sensory stimulation to fit its model.

A novel consequence of PP is that it entirely reverses the traditional conception of how processing occurs within the brain. The brain is usually thought to process sensory inputs in a *bottom-up* manner [Marr, 1982]. Bottom-up views assume that brain processing begins at the periphery of the sensory system, with the information registered at this point being extremely simple. For example, visual processing will be considered to begin at the retina with only extremely simple information registered at this point (concerning, for example, line orientation). This information is thought to be encoded into representations, and these representations are thought to be passed up the cortical hierarchy and enriched with additional representational content. For example, visual representations will be passed up the visual hierarchy with information added at various stages that serves to make the visual representation more complex (colour information is encoded at V4, motion information at V5, and so on). In short, the bottom-up model explains perception to be constituted by the brain using its representational capacities to pass the (initially impoverished) information registered at its peripheries up the cortical hierarchy, in order to arrive at the perceptual states that we experience.

Figure Three: Bottom-Up Model

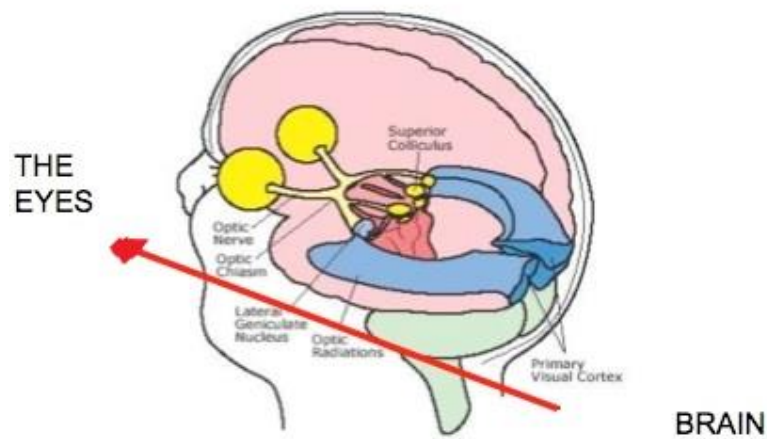


[Skillen, 2009]

PP entirely reverses this traditional view because it takes brain processing to occur in a *top-down* manner. According to PP, the brain does not receive simple information at the sensory peripheries and make it more complex as it is passed *up* the cortical hierarchy. Rather, the brain creates a complex representation (the winning hypothesis) and this representation is projected *down* the cortical hierarchy, serving as a constraint on cortical processing.

Consider the following diagram:

Figure Four: Top-down processing



[Skillen, 2009, *modified*]

Consequently, PP requires a revised conception of bottom-up neural processing. The traditional bottom-up view understands bottom-up signaling to involve the passing of sensory information received from the environment up the cortical hierarchy. On this view, bottom-up signals play the largest role in determining the content of perceptual representations. PP, however, takes bottom-up processing to signal errors in the brain's top-down predictions. These *error-signals* signal that there has been a divergence between the predicted sensory input and the actual sensory input received. The content of perceptual representations in PP is therefore mostly determined by top-down processing.

The PP reconceptualization of what bottom-up signals are accounts for an anatomical fact that has long puzzled psychologists who accept the bottom-up model— the brain has many more top-down connections than bottom-up ones [Callaway, 1999;



Felleman & Van Essen, 1991]. This anatomical fact is difficult to account for on bottom-up theories because they predict the reverse. On these theories it would be expected that there would be many more bottom-up connections in the brain because brain processing is largely a bottom-up affair. PP, however, proposes that brain processing is reliant mostly on top-down connections. Consequently, it would be expected on this theory that there would be many more top-down anatomical connections in the brain.

### ***§1.2: How Does PP Propose to Explain the Sub-personal, Personal, and Conscious Levels of Explanation?***

Cognitivist theories tend to explain the inter-relation between the sub-personal, personal, and conscious levels of explanation via mechanistic decomposition. In this sub-section I briefly explain what mechanistic explanations in cognitive science are, show that cognitivist mechanistic explanations make indispensable use of representation, and then provide two examples of mechanistic accounts of PP theories of perception which make indispensable use of representation. In doing so, I show that cognitivist accounts (such as PP) make indispensable use of representation in their attempts to bridge and explain the inter-relation between the sub-personal, personal, and conscious levels of explanation.

#### ***§1.2.1: Mechanistic Explanations In Cognitive Science***

Cognitivists tend to favour *mechanistic* explanations of the relation between the sub-personal, personal, and conscious levels of explanation. There is a vast literature on mechanistic explanations in cognitive science (and indeed in science more generally) [Craver

& Tabery, 2016], but here I will focus on providing a brief and representative outline of mechanistic explanation [Bechtel, 2009; Gladziejewski, 2015; Milkowski, 2016].

Mechanistic explanations, as the name would suggest, proceed by providing an explanation of the specific mechanism responsible for generating the phenomenon under study. In order to provide a mechanistic explanation, the following steps must be followed—first, a specific phenomenon for study is identified. Then, the mechanism underlying this phenomenon is identified and modelled. Finally, theorists must explain how the components parts of that mechanism work together in order to allow for its proper functioning. If these steps are successfully followed, then one will have arrived at a mechanistic explanation of the phenomenon in question.

Mechanistic explanations in cognitive science proceed by first identifying a cognitive capacity to be explained mechanistically. In order to explain the capacity in question, a mechanism responsible for its instantiation must be proposed. Finally, the cognitive scientist must explain how the component parts of the mechanism in question work together in order to enable the capacity at issue. Consider the example of spatial navigation in rats. Scientists often propose the hippocampus as the mechanism responsible for spatial navigation [O’Keefe & Nadel, 1978]. In order to substantiate this proposal, in the case of rat spatial navigation, the scientist must explain how the various component parts of the rat hippocampus combine in order to enable successful spatial navigation. The success of this mechanistic model will in turn be determined by the extent to which it successfully explains rat spatial navigation.

Mechanistic explanations of cognition proposed within cognitivist cognitive science make indispensable use of representation:

Mechanisms in cognitive science, in contrast, are proposed to explain *cognitive* activities such as memory retrieval or problem solving by performing operations on *representations* that carry information about objects, events, and circumstances currently or previously encountered. Operating on representations is different than merely moving or transforming physical substances, in that representations serve an informational function: they relate a vehicle (the form of the representation) to a content (what it is about). [Bechtel, 2009, p. 553-554, *italics in original*]

These representational mechanistic explanations of cognition proceed as detailed above (specify cognitive capacity, propose components of cognitive mechanism, and so on).

Importantly, however, they require that “representation” is itself an important component of the mechanism being modelled:

I propose that representational explanation in cognitive science is also a sort of mechanistic explanation. That is, cognitive scientists who postulate internal representations are involved in the project of searching for mechanisms that underlie cognitive phenomena. Since representations are usually understood as internal states or structures somehow located inside the cognitive system— that is, they are understood as component parts of the cognitive system— I think it is natural to treat representational explanations as mechanistic. [Gładziejewski, 2016, p. 66]

Thus, the mechanism (or a component part of that mechanism) is thought to play a representational role within the wider cognitive system of which it is a part.

Cognitivists, therefore, use mechanistic explanations in order to bridge the divide between sub-personal, personal, and conscious levels of explanation— a given cognitive

phenomenon is specified, and then a mechanistic account of this phenomenon is provided which explains how it could be implemented within the brain. These mechanistic accounts make indispensable use of representation and, therefore, the concept of “representation” plays a crucial role in cognitivist attempts to explain the inter-relation between sub-personal, personal, and conscious levels of explanation.

### **§1.2.2: Perceptual Mechanisms, PP, and Representation**

It is commonly assumed that mechanistic explanations of PP theories of perception indispensably require representation [Clark, 2013; 2016; Hohwy, 2013]. In this section I will outline two specific proposals as to why the mechanistic implementation of PP theories of perception requires representation.<sup>16</sup>

The first proposal I will consider is provided by Pawel Gladziejewski [2016].

Gladziejewski structures his proposal around William Ramsey’s *job-description challenge*. In his 2009 *Representation Reconsidered*, Ramsey argues that a given cognitive mechanism should be assigned a representational status only if it meets the job-description challenge. Ramsey notes that almost any mechanism can trivially be described in representational or non-representational terms. Consequently, in order to avoid trivialising the issue of whether a given cognitive mechanism is representational, he argues that the mechanism in question should play a recognisably representational role in the cognitive system in question in

---

<sup>16</sup>Although some theorists are sceptical that PP models can be explained mechanistically [Milkowski, 2016], it is generally agreed by proponents of PP that neural implementations of the theory must be found eventually. If not, the theory “fails as a distinctive empirical account” [Clark, 2013, p. 200]. At this point in time, the project of providing such neural implementations of PP models is a relatively new but active area of research. The two models I will describe are consequently relatively light on neuroanatomical detail. However, this lack of neuroanatomical detail simply reflects the current state of research. The important point, for present purposes, is that the proposed mechanistic implementations of PP indispensably require representation.

order to pass the job-description challenge. Ramsey submits that one effective way to determine whether a given mechanism plays a representational role is to apply the *compare-to-prototype strategy*. This strategy involves describing a prototypical representation (such as a map), and then determining whether or not the mechanism under discussion plays a similar role to the prototypical representation (for example, does the mechanism play a recognisably map-like role within the cognitive system). If the mechanism in question does fulfil a functional role similar to the prototypical representation in question, then the mechanism can be described in terms of representation.<sup>17</sup>

Gladziejewski applies the job-description challenge to the mechanistic implementation of PP, and he concludes that the mechanistic implementation of PP does require representation:

I claim that representations as PCT [predictive coding theory] sees them earn their status because the functional roles they play non-trivially resemble the roles played by *cartographic maps*...The representational posits of PCT fit the same job-description as cartographic maps. [Gladziejewski, 2016, p. 566]

Gladziejewski defines maps as (1) *structural representations* which (2) guide action, (3) are detachable, and (4) can misrepresent. Maps are considered to be structural representations because they represent by possessing a structure which is isomorphic with the place being mapped. For example, black lines on a bus route map are structurally isomorphic to the actual routes taken by certain buses and represent these routes for this reason. Maps are often used for the guidance of action. If I wish to get a certain bus, I will consult the bus

---

<sup>17</sup>I discuss Ramsey's job-description challenge in more detail in chapter five of the thesis.

route map and subsequently make my way to the location of the bus stop indicated on the map. They are considered detachable because they can represent a location or area even when not themselves spatio-temporally located in that area. For example, I can consult the bus route map of another country from the comfort of my home. Finally, maps can misrepresent. If my bus route map is out of date, or if there has been a mistake during its production, then the map may depict a certain bus route which does not exist.

Gladziejewski uses maps as proto-typical representations in his own application of the compare-to-prototype strategy. He argues that the *internal generative models*<sup>18</sup> posited by PP play the role of probabilistic maps which function as (1) structural representations that (2) guide action, (3) are detachable, and (4) can misrepresent. Gladziejewski argues that PP representations are structural representations because they recapitulate “the causal–probabilistic structure of the environment” [Gladziejewski, 2016, p. 571]. He claims that they fulfil condition (2) because *active inference* requires that prediction error-minimisation results in action, and so internal generative models are therefore used by the PP system to guide action. Then, he argues that internal generative models meet condition (3) because the PP organism acts on the basis of its internal generative model of the world, and not the world itself. (The internal generative model is supposed to be, after all, a model of the world— it is the brain’s best guess as to what is out there.)<sup>19</sup> Finally, Gladziejewski concludes that internal generative models fulfil condition (4) because they can misrepresent in one of two ways: they can misrepresent because the dominant hypothesis itself

---

<sup>18</sup>This is a technical term used to describe the top-down hypotheses which constrain and determine cortical processing in lower levels that were discussed in §1.1.

<sup>19</sup> Gladziejewski accepts that the internal generative model may only be partially detachable. Because I am only outlining sample PP accounts in order to show why they are generally taken to require *representational* mechanistic explanations, I will not delve into this point here.

inaccurately models external events; or, because the dominant hypothesis accurately models external events but inaccurately models the organism's own position amongst them. For example, misrepresentation can occur when the brain includes a chair in its model of the world when no chair is present, or it can occur when the brain correctly models a chair but incorrectly takes the organism to be sitting on that chair. Consequently, Gladziejewski concludes that internal generative models fulfil the same functional role as maps (which are proto-typical representations). Therefore, they meet the job-description challenge and so deserve to be described in terms of representation.

Anil Seth is another PP theorist who has argued that a mechanistic implementation of PP will require representation [2014]. He argues that the theoretical apparatus of PP can be used to operationalise the concept of "sensorimotor knowledge". According to Seth, active inference provides an operationalisation of sensorimotor contingencies because it is essentially concerned with the law-like relation between sensation and movement. However, he claims that *mastery* or *knowledge* of sensorimotor contingencies can only be operationalised by:

HGMs [hierarchical generative models] encoding *conditional* aspects of SMCs [sensorimotor contingencies...that] would incorporate explicitly *counterfactual* probabilistic models of the behavior of hidden causes of fictive sensory signals (and their precisions) given particular actions. [Seth, 2014, p. 104, *italics in original*].

In chapter two I explained that SE takes perception to occur only when an organism possesses and exercises sensorimotor knowledge (knowledge of the law-like relation between sensory stimulation and movement). An important aspect of perception concerns the fact that, when we perceive a given object, we perceive more than that which is (strictly

speaking) present to our sensory system. For example, when I perceive a tomato I perceive it as a three-dimensional object even though I do not (at the present moment) have perceptual access to its back and sides. Alva Noë labels this phenomenon “*perceptual presence*”, but it is more commonly known as “*amodal completion*”. Seth suggests that perceptual presence can be explained via a hierarchical<sup>20</sup> generative model which explicitly encodes counter-factual aspects of perceptual experience. In the case of the tomato, he suggests that we perceive it as a three-dimensional object because we possess an internal generative model which encodes not only the aspects of the tomato directly facing us, but also aspects of the tomato which would be perceptually experienced if we were to engage in a particular action (for example, the tomato’s back, which would come into view were we to perform certain movements).

Seth takes his own PP account to improve upon SE because it can provide a neural implementation of sensorimotor knowledge:

However, unlike the PP view, and despite the attention the theory has received over many years, possible neural or mechanistic implementations [of SE] remain unspecified. [Seth, 2014, pp. 102-103]

Importantly, in providing his own mechanistic account of PP Seth makes indispensable use of representation. As Seth himself notes, his theory “departs from standard versions of sensorimotor theory which, in contrast to PP, are explicitly non-inferential” [*ibid.* p. 105]. Seth accepts a cognitivist view of perception, and his invocation of explicit encoding of counter-factual aspects of perception in order to account for perceptual presence requires

---

<sup>20</sup>PP models are often described as “hierarchical” because they involve the passing down of predictions from higher to lower neural areas.



representation. Once more, we can see that mechanistic explanations of perception within the PP framework make indispensable use of representation.

PP is a cognitivist theory of perception. Its proponents do advocate the method of mechanistic decomposition as a means of explaining the inter-relation between sub-personal, personal, and conscious levels of explanation. Where perception is concerned, proponents of PP argue that the brain uses sub-personal *representations* and that its use of these internal representations allows for and explains personal level and conscious perception [Clark, 2013; 2016; Hohwy, 2013, ch. 8; 2014; Gładziejewski, 2016; Rescorla, 2015; Seth, 2014]. The key point to be taken from this discussion is that extant PP models of perception *require* representation. It is this requirement— that mechanistic PP explanations of perception *require* representation— which leads to the *hard problem of perceptual consciousness* (or so I shall argue).

## **§2: Cognitivism and the ‘Hard Problem of Perceptual Consciousness’**

In this section I am going to argue that cognitivist accounts rely on a conceptual distinction which allows for the ‘hard problem of perceptual consciousness’. I begin by explaining what the ‘hard problem of perceptual consciousness’ is, and arguing that the most popular response to this problem fails. Then, I explain why the cognitivist rejection of the tenets of *activity* and *knowledge-how* allows for the ‘hard problem’ and apply this explanation to PP. Consequently, I conclude that PP can only account for the conscious level of description by giving rise to the ‘hard problem of perceptual consciousness’.

### **§2.1: What is the ‘Hard Problem of Perceptual Consciousness’?**

The *hard problem of consciousness* is the problem of explaining why conscious experience exists at all. There is ‘something it is like’ to experience but it seems to be extremely difficult to explain this ‘something it is like’ in functional or mechanical terms. As such, it seems mysterious why there would be ‘something it is like’ to experience at all. When applied to the case of perception, the ‘hard problem’ concerns why there is ‘something it is like’ to perceive. This problem has a venerable tradition in philosophy, having appeared in a number of guises throughout history. I will here focus on one particular version of the argument— the modal argument. This argument has been advanced by a number of theorists (for example, Descartes [1641]; Kripke [1980]). In this chapter I will focus on its most popular contemporary construal— David Chalmers’ *zombie argument* [1996].

Chalmers proposes the following thought experiment— imagine a universe (the *zombie-universe*) which is physically identical to our own and yet in this universe conscious perceptual experience does not exist.<sup>21</sup> Although zombie-humans carry out the exact same perceptual activities in zombie-universe as we do in our own, there is nothing it is like for them to do so. Zombie-humans will, for example, visually perceive objects but there will be nothing it is like for them to do so. If a zombie-universe is conceptually coherent (in that there is nothing contradictory or nonsensical about philosophical zombies) then it follows that zombies are conceivable. If zombies are conceivable then, according to Chalmers, they are metaphysically possible. The metaphysical possibility of zombies entails that there is only a contingent relation between physical states and phenomenal states. In our world

---

<sup>21</sup> The zombie argument is generally concerned with conscious experience *tout court*. Because the topic of this thesis is conscious perception, I will therefore be outlining a perception-specific version of the argument.

phenomenal states are caused by or correlated with physical states. However, there is no necessary connection between physical states and phenomenal ones, and any such connection is purely contingent.<sup>22</sup> The upshot is that if zombies are conceivable, then physical states alone cannot account for the existence of conscious perception. Therefore, we cannot explain conscious perception physically.

### ***§2.1.2: The Phenomenal Concept Strategy***

There are general metaphysical concerns with Chalmers' move from conceivability to possibility which I will not deal with here because doing so would go far beyond the scope of this thesis.<sup>23</sup> Instead of dealing with these general metaphysical concerns, I will focus on one instance of these concerns which is specific to the zombie argument itself— the *“phenomenal concept strategy”*.

*Type-B materialists* accept that zombies are conceivable but they reject Chalmers' move from conceivability to possibility. Consequently, type-B materialism involves attempting to explain how our intuitions that zombies are conceivable could be entirely compatible with the truth of physicalism. The phenomenal concept strategy is a type-B materialist position. Chalmers summarises the phenomenal concept strategy as follows:

---

<sup>22</sup> Consider that although I did actually write this chapter, this just happens to be a contingent fact about myself that I did so. There would be nothing contradictory about a world in which I did not write this chapter. For Chalmers, the same point applies to phenomenality— although phenomenal states happen to ‘come along’ with physical states in this world there is no reason they have to. The relation between physical and phenomenal states is just as contingent as the relation between my identity and my having written this chapter.

<sup>23</sup> The interested reader should consult [Chalmers, 2010, chs. 5-6] for a comprehensive defence of this move.

Those who react in this way agree that there is an explanatory gap, but they hold that it stems from the way we *think* about consciousness. In particular, this view locates the gap in the relationship between our concepts of physical processes and our concepts of consciousness, rather than in the relationship between physical processes and consciousness themselves. [Chalmers, 2010, p. 305, *italics in original*]

The phenomenal concept strategy requires accepting dualism about the *concepts* of physical and phenomenal whilst rejecting dualism about their *metaphysical status*. Accepting dualism about our concepts of physical and phenomenal allows one to explain why it is that zombies are conceivable. Rejecting dualism about the metaphysical status of these concepts, however, allows one to reject Chalmers' move from the conceivability of zombies to their metaphysical possibility. Consequently, proponents of the phenomenal concept strategy argue that we are tricked into believing that there is a metaphysical 'hard problem of perceptual consciousness' because our concepts of the physical and the phenomenal have misled us.

Chalmers rejects the phenomenal concept strategy by arguing that it falls onto the horns of a dilemma. Either the proponent of this strategy will accept:

a 'thick' account, in which the relevant features of phenomenal concepts are not physically explainable (although they may explain our epistemic situation), or a 'thin' account, in which the relevant features of phenomenal concepts do not explain our epistemic situation (although they may be physically explainable). [Chalmers, 2010, p. 332].

The phenomenal concept strategy falls onto the horns of a dilemma because ‘thick’ accounts of phenomenal concepts will give rise to their own explanatory gap problem, whereas ‘thin’ accounts will not do enough to motivate the idea that there is a conceptual distinction between the physical and phenomenal in the first place.

Chalmers defines ‘thick’ accounts as follows:

[‘Thick’ accounts] build in special epistemic features of phenomenal concepts: the idea that phenomenal states present themselves to subjects in especially direct ways, or the idea that simply having a phenomenal state enables a certain sort of knowledge of the state, or the idea that the state itself constitutes evidence for the state. If we build in such features, then we may be able to explain many aspects of our distinctive epistemic situation with respect to consciousness. [Chalmers, 2010, p. 335-336]

Chalmers contends that, although these accounts of our phenomenal concepts do explain why we find zombies conceivable, they cannot do so in a physically satisfactory manner:

If these features are powerful enough to distinguish our epistemic situation from that of a zombie, then they will themselves pose as much of an explanatory gap as does consciousness itself. [Chalmers, 2010, p.336]

If one accepts a ‘thick’ account, one will take there to be a qualitative difference between humans and zombie-humans which can be discerned from the first-person perspective. If there is such a difference, however, then there will be an obvious experiential difference between humans and zombie-humans. Rather than having to explain consciousness, the proponent of ‘thick’ accounts must instead explain the epistemic properties of humans

which distinguish them from zombie-humans. Given that this epistemic difference distinguishes humans from zombie-humans, positing it will not help the phenomenal concept strategist in their aim of explaining the conceivability of zombies whilst rejecting their metaphysical possibility. 'Thick' accounts give rise to their own metaphysical explanatory gap problem.

'Thin' accounts are capable of explaining phenomenal concepts in a physically satisfactory manner. Such accounts will provide a physical explanation as to how we come to possess the concepts of "physical" and "phenomenal" and explain why these concepts appear to refer to different things. For example, it could be argued that the areas in the functional architecture of the brain which are concerned with phenomenal experience and language processing operate independently of one another. Consequently, although the physical and phenomenal are in fact one and the same thing, the isolation of the functional processes underlying phenomenal experience and linguistic processing makes it appear as if they are different.

'Thin' accounts do explain how we could come to possess conceptual dualism about the concepts "physical" and "phenomenal". However, they are not going to be helpful if their proponent intends to defend a type-B materialist position. Such accounts do nothing to explain why we should resist the move from the conceivability of zombies to their metaphysical possibility, because they require the inconceivability of zombies. If our distinct concepts of "physical" and "phenomenal" are predicated on the nature of our cognitive architecture and this architecture is physical, then it follows that zombies should be considered inconceivable. A 'thin' account of phenomenal concepts will therefore not allow

for the conceivability of zombies. It will require that there is no difference between humans in this world and humans in zombie-world.

In this section we have seen that the most popular type-B materialist position, known as the phenomenal concept strategy, fails. Proponents of this position end up either: accepting a ‘thick’ account of phenomenal concepts and therefore facing the ‘hard problem of perceptual consciousness’ all over again; or, they will accept a ‘thin’ account which requires the inconceivability of zombies. Consequently, in present circumstances, the phenomenal concept strategy will not assist the physicalist who intends to avoid the zombie problem by accepting the conceivability of zombies whilst rejecting their metaphysical possibility. I will therefore assume in what follows that if one is to reject the zombie argument, one must explain why zombies are inconceivable in a phenomenologically satisfactory manner.

### ***§2.3: Cognitivism and the ‘Hard Problem of Perceptual Consciousness’***

In chapter one I explained that cognitivist positions explicitly reject the tenets of *activity* and *knowledge-how*. Accordingly, cognitivist theories of conscious perception are predicated on the following two claims:

1. *~Activity*: perception is not constituted by action. (Perception may inform an organism’s actions in the world, but it is not itself constituted by actions).
2. *~Knowledge-How*: perception is constitutively representational.

I am going to argue that acceptance of (1) and (2) inevitably leads the cognitivist to the ‘hard problem of perceptual consciousness’ and will apply this argument to the case of PP. My

argument proceeds in two steps— first, I explain why rejecting *knowledge-how* allows for a conceptual distinction between physical and phenomenal states. Then, in step two, I argue that rejecting *activity* results in an inability (for the cognitivist) to require a necessary connection between physical states and phenomenal ones. I will therefore conclude that cognitivist positions should not be used to explain consciousness because they give rise to the ‘hard problem of perceptual consciousness’.

*Step One: Perception is Representational*

1.  $\sim\text{Knowledge-How} \rightarrow$  Perception is representational
  2. Mental representation is non-conscious  $\rightarrow$  There can be mental representation without consciousness
  3. There can be mental representation without consciousness  $\rightarrow$  There is no necessary connection between mental representation and consciousness
- Therefore
4.  $\sim\text{Knowledge-How} \rightarrow$  There is a conceptual distinction between physical representational states and conscious experience.

If one rejects *knowledge-how* then one must explain perception in terms of mental representation (1) (as explained in chapter one). It is generally agreed that one can have mental representation in the absence of consciousness (2).<sup>24</sup> If one can have mental

---

<sup>24</sup> This step could be rejected by accepting the position known as “*Phenomenal Intentionality*”. On this view representational content is essentially derivative from consciousness, and so consciousness is a pre-requisite if representation is to exist. As such, there will be a necessary connection between representational states and phenomenal experience (because a world without phenomenality is also a world without representation). I am inclined to reject the phenomenal intentionality programme for two reasons:

1. It requires one accept an anti-naturalist theory of ‘original content’ [Hutto & Myin, 2013]
2. It accepts a view of content which is incompatible with the methods and practise of cognitive science [Egan, 2013].

The whole point of providing a physically satisfactory explanation of consciousness is to bring the phenomenon into the realm of natural science. Accepting phenomenal intentionality requires rejection of naturalism (this is



representation which is not conscious, then it follows that there is no necessary connection between mental representation and consciousness (3). If there is no such necessary connection, then there is a conceptual distinction between physical representational states and phenomenal experience (4). Consequently, we arrive at the conclusion that there is nothing contradictory or incoherent in the idea of there being a mental representation which is not conscious.

Let us use the example of PP to fill in this argument. PP rejects *knowledge-how* and it explains perception in terms of representation (as explained in §1, *cf.* [Clark, 2013; 2016; Gladziejewski, 2016; Hohwy, 2013, ch. 8; 2014; Rescorla, 2015; Seth, 2014]). PP accepts that one can have mental representation in the absence of consciousness, and should therefore be led to the conclusion that there is a conceptual distinction between physical representational states and phenomenal experience. This is indeed the conclusion that PP theorists arrive at:

Conscious perception is the upshot of unconscious perceptual inference. We are not consciously engaging in Bayesian updating of our priors in the light of new evidence, nor of the way sensory input is predicted and then attended. What is conscious is the result of the inference— the conclusion. That is, conscious perception is determined by the hypotheses about the world that best predicts input and thereby gets the highest posterior probability. [Hohwy, 2013, p. 201]

---

the take home point of problems (1) and (2)). Therefore, in the present context I do not consider the position appealing.

At the heart of that [PP] story...is the kind of automatically deployed, deeply probabilistic, non-conscious guessing that occurs as part of the complex neural processing routines that underpin and unify perception and action. [Clark, 2016, p. 2]

By definition, unconscious inference is not conscious. According to PP, unconscious representational states are not only conceptually possible, they in fact nomologically exist. Therefore, PP does not take there to be a necessary connection between mental representational states and phenomenal experience. Consequently, for PP there is no incoherence in the idea of an inferential and brain-based mechanism representing the world, with the representation produced failing to result in conscious experience.

*Step Two: Perception is not constituted by Activity*

1. Perception is representational → Perception is not constituted by action
2. Perception is not constituted by action → There is no necessary connection between perception and action

If perception is constituted by representational content, then it will not be constituted by an organism's actions— provided there is representational content there will be perceptual experience. Perceptual experience will be constituted by the content of a given state, and a given state's content is not constituted by action or behaviour (1). On the cognitivist view the conscious mind is brain-bound because its representational vehicles are brain-bound. Action itself will therefore not be constitutive of consciousness, because action occurs outside of the brain (it is for this reason that cognitivist theories avoid the problem of verificationism). There can be perceptual experience in the absence of behaviour, and so

cognitivism allows for a conceptual distinction between perceptual experience and behaviour (2).

PP does take the representational vehicles of consciousness to be brain-bound [Clark, 2009; 2012; 2016; Hohwy, 2013; 2014; Seth, 2014]. Consequently, on this framework an organism's behaviour will not be considered constitutive of perceptual experience (PP allows only for a causal relation between perceptual experience and behaviour). Therefore, according to PP, there is a conceptual distinction between perceptual experience and an organism's behaviour.

In step one of the argument, it is concluded that there is no necessary conceptual connection between representation and phenomenality. In step two, it is concluded that there is no necessary conceptual connection between behaviour and phenomenality. Cognitivist theories are not relational theories, and thus representation and/or behaviour exhaust the set of physical things which could be relevant for explaining conscious perception (in these theories). For cognitivists there is no necessary connection between representation and/or behaviour and consciousness, and so cognitivists must arrive at the conclusion that there is a conceptual distinction between the physical and the phenomenal. Therefore, for the cognitivist, perceptual-zombies are conceivable.

This argument was applied to PP. In step one, we saw that PP admits a conceptual distinction between representation and consciousness. In step two, we saw that PP admits a conceptual distinction between behaviour and consciousness. Because PP does not allow for the environment to play a constitutive role in perceptual experience, representation and behaviour exhaust the set of physical things which could be nomologically related to conscious perception. Consequently, PP must allow for a conceptual distinction between

physical and phenomenal states and therefore faces the ‘hard problem of perceptual consciousness’.

PP theorists are quite forthright in admitting that this is a consequence of their view:

[T]his is not intended as a proposal that can explain *why* perceptual states are phenomenally conscious rather than not. It is merely a proposal that describes the states that *are* conscious; they are those representations of the world that currently best predict sensory input. I am not here intending to touch upon the *metaphysical* mind-body problem— the ‘hard problem’ of consciousness. [Hohwy, 2013, pp. 202, *italics in original*; cf. Clark, 2016, ch. 7, ft. 26]

In summary, I have argued that PP cannot explain conscious perception without admitting a conceptual distinction which results in the ‘hard problem of perceptual consciousness’.

Naturalistic theories should explain conscious perception without facing the ‘hard problem’.

PP does not avoid this problem, and so does not provide one with an adequate naturalistic theory of conscious perception.

### **§3: Dissolving the ‘Hard Problem of Perceptual Consciousness’**

In this section I will argue that one can avoid the conceptual coherence of the ‘hard problem of perceptual consciousness’ by accepting the tenets of *activity* and *knowledge-how*. I begin by outlining Gilbert Ryle’s dissolution of the ‘hard problem’, and explaining that it is generally rejected because it is phenomenologically implausible. Then, I will argue that RSE can accept Ryle’s dissolution of the ‘hard problem of perceptual consciousness’ whilst providing a phenomenologically plausible account of conscious perception. Consequently, I

conclude that RSE can provide a phenomenologically plausible explanation of conscious perception without facing the ‘hard problem of perceptual consciousness’.

### ***§3.1: Ryle’s Dissolution of the Problem and its Verificationist Consequences***

Gilbert Ryle famously provided an influential dissolution of the problem of consciousness [1949/2000, ch. 1]. He argued that the very idea of a philosophical zombie is based on conceptual confusion— zombies are inconceivable because one cannot dissociate consciousness and action in the manner required by the zombie argument.<sup>25</sup> Ryle’s position is capable of avoiding the ‘hard problem of perceptual consciousness’ because it takes conscious perceptual experience to be constituted by an organism’s dispositions to engage in perceptual behaviour [1949/2000, ch. 7]. For Ryle, ‘what it is like’ to experience perceptually is constituted by a collection of perceptual behaviours. As such, explaining ‘what it is like’ to consciously perceive requires no additional ingredient or naturalistically mysterious properties over and above action, and so philosophical zombies are inconceivable. On Ryle’s view of conscious perception, the very idea of an agent physically identical to us and yet lacking in phenomenal experience is incoherent.

Ryle’s dissolution of the ‘hard problem of consciousness’ is generally rejected for having verificationist consequences.<sup>26</sup> His position is problematic because his verificationist tendencies lead him to conflate the effects of a particular mental state with the mental

---

<sup>25</sup>Given that the zombie argument was formulated some twenty years after Ryle’s death, it is somewhat anachronistic to express my point in this manner. Consequently, it would be more accurate to state that Ryle’s position entails that the conceivability of zombies is based on conceptual confusion.

<sup>26</sup> Although Ryle is often labelled a verificationist, it has been argued that categorising him in this manner is unfair (see [Tanney, 2015] for a recent defence of this point). Because I do not intend to engage in Ryle scholarship here I will assume that the dominant interpretation of Ryle, upon which his position is considered to result in verificationism, is correct (for a recent argument to this effect, see [Stanley, 2011, ch. 1, §1]).

state itself. Consequently, his theory is unable to account for ‘what it is like’ to experience. Consider the example of visual perception. Ryle argues that visual perception is constituted by an organism’s visual behaviour. This is problematic, however, because we can conceive of a being who visually perceives and yet which does not exhibit any behavioural evidence of having done so. For example, a person could be visually perceiving and yet providing no behavioural evidence of doing so (say, because they are making a philosophical point) [*cf.* Putnam, 1975, ch. 16]. In fact, people have allowed themselves to be injected with drugs which completely paralyse them (temporarily), and it has been demonstrated that these people perceive despite being unable to perform any kind of perceptual behaviour [Aizawa, 2015]. The problem, therefore, with Ryle’s explanation of conscious perception is that he conflates the effects of conscious perception (the perceiver’s behaviour) with conscious perception itself.

Ryle’s dissolution of the ‘hard problem of perceptual consciousness’ is usually rejected because it cannot provide a phenomenologically satisfactory account of ‘what it is like’ to experience perceptually. Consequently, Ryle can deny the conceivability of perceptual zombies only because he denies the manifest facts of conscious perceptual experience [*cf.* Chalmers, 1996; Levine, 2004].<sup>27</sup>

---

<sup>27</sup> This response is extremely popular— philosophers generally agree that we should not deny the conceivability of zombies, with only 16% of professional philosophers taking zombies to be inconceivable [Bourget & Chalmers, 2014].

### ***§3.2: Dynamic Views and Verificationism***

I am going to argue that RSE can join Ryle in rejecting the conceivability of perceptual zombies because it accepts the tenets of *activity* and *knowledge-how*. Because conscious perception is constituted by actions (performed attentively) which themselves are explained without representation, RSE does not allow for a conceptual distinction between physical states and phenomenal ones. Consequently, in order to explain how RSE can provide a phenomenologically satisfactory dissolution of the 'hard problem of perceptual consciousness', I must explain why it avoids the problem of verificationism.

Although Ryle's theory of conscious perception does have verificationist consequences, his dissolution of the 'hard problem of perceptual consciousness' does not prescribe a position with verificationist consequences. Taking action to be constitutive of perception does not in and of itself lead to verificationism. Rather, Ryle is led to verificationism because he accepts both the thesis that perception is constituted by behaviour and a *linear input-output* view of perception (or so I will argue). Susan Hurley has argued that, provided one rejects a linear theory and instead accepts a *dynamic* theory, one can take conscious perception to be inextricably linked with action and yet avoid the problem of verificationism [Hurley, 1998, pp. 420-422; 2001].

The view of mentality which Susan Hurley labels as the *linear input-output* view can be described as follows:

The organism receives perceptual input → Cognition occurs in the brain → A behavioural output is produced

On this picture of the mind, mentality is thought to begin at the sensory periphery of the organism (determined by where perceptual information is registered). After perceptual input has been received, cognitive operations will be performed on it. These will result in the formulation of a motor plan for action that is sent to the motor system and implemented. The most important thing to note is that, on the traditional input-output picture, perception and action are taken to be wholly isolated and independent aspects of mentality.

Hurley argues that behaviourists like Ryle are not forced into verificationism because they take action to be inextricably inter-related with perception. Rather, they are forced into verificationism because they take action to be inextricability inter-related with perception whilst accepting an input-output view of mind:

Behaviourism and related views take as their point of departure a framework in which action is seen as an effect of perception. Feedback from output to input is ignored or considered to be a mere complication, not of the essence...[T]he primitive and more sophisticated versions of the framework have in common a linear or one-way view of the primary causal flows...

From a point of departure with this linear and one-way character, behaviourist views make a constitutivising move. They move from seeing perceptions and actions as separate, causally and instrumentally related events, to seeing action as constitutive of perception. Given the point of departure, this move seems to collapse perceptual experience into its effects, to reduce perceptual experience to what is merely the evidence for it in behaviour. So the objection arises that behaviourism is a form of verificationism. [Hurley, 2001, p. 17]



In short, if one takes action to be inextricably inter-related with perception and one accepts an input-output view of mind, then one will be forced to accept a verificationist position.

Hurley argues, however, that if one takes action to be inextricably inter-related with perception and rejects an input-output view of mind, one can avoid the problem of verificationism:

[T]he charge of verificationism here presupposes the prior one-way framework, the point of departure that sees actions as effects, output. To test that claim, compare a constitutivising move made within a different framework. Suppose our point of departure sees the causal flows as essentially circular, regards actions to be as much causes as effects of perceptions, and sees both as emerging from a complex dynamic feedback system. Suppose we now consider the possibility that perceptions and actions are not necessarily separate, causally related events, but that action can be constitutively related to perception...

In this context, the constitutivising move no longer has a verificationist character. It is not the constitutivising move itself that leaves behaviourism open to the charge of verificationism. It is rather making this move against the background of the linear, one-way framework. If we fail to recognise that behaviourism is not the only way to give action a constitutive relationship to perception, our sense of the range of philosophical options will be distorted. [*ibid.* pp. 17-18]

Dynamic views take perception and action to be so intimately intertwined that it is not possible to make any sort of principled distinction between the two. As such, they completely reject the linear input-output idea of perception and action as autonomous from one another, and instead view action and perception as inseparable. Andy Clark [1999] uses

the helpful metaphor of *Escher Spaghetti* to get this point across, stating that “[mentality is] not just multiple criss-crossing strands (ordinary spaghetti), but strands whose ends feed back into their own (and others) beginnings, making 'input' and 'output', and 'early' and 'late' into imprecise and misleading visions of complex recurrent and re-entrant dynamics” [*ibid.* p. 11].

According to Hurley, one can take action to be inextricably inter-related with perception and yet avoid verificationism by replacing a linear input-output view of mind with a dynamic one. However, emphasising the dynamic aspects of perception alone will not (as yet) allow for a dissolution of the ‘hard problem of perceptual consciousness’. Hurley only argues that action is inextricably inter-related with perception, such that potential perceivers must also be potential actors (*cf.* [Ward, 2015; 2016]). She does not, however, make the stronger claim that perception is *entirely constituted by* action. It is only by making this stronger claim that one can avail of Ryle’s dissolution of the ‘hard problem of perceptual consciousness’.

Direct realist theories of conscious perception allow for factors external to the brain to be constitutive of conscious experience. If one allows for such factors to play a constitutive role in experience, and one accepts a dynamic view of mentality, then one can extend the constitutive role of action in perception beyond agential factors and into the environment itself [*cf.* Clark & Chalmers, 1998]. In other words, both environmental and agential action should be considered to be constitutive of consciousness. This extended conception of the constitutive role of action in perception still requires that if no action occurs, then there will be no perception (recall that this was the problem with Ryle’s theory— there appeared to be perception even when no action took place). However,

rather than being problematic for dynamic views which allow environmental action to play a constitutive role in perception, this turns out to be a positive prediction of such views. It turns out empirically that if there is no movement *whatsoever* then there will be no perceptual experience. This phenomenon is known as *sensory fatigue* and is accurately summarised (in the case of vision) as follows: “Non-constancy of retinal stimulation is necessary for visual processing.” [Aizawa, 2015]. If there is no movement on behalf of the organism *or* on behalf of the object being perceived, then there will be no perceptual experience. For example, if one’s eyes are focused on a visual stimulus and the stimulus itself does not move or change in some way, it will quite literally fade from view and no longer be perceived [Martinez-Conde, Macknik, & Hubel, 2004]. Therefore, we can see that if no action *at all* occurs there will be no experience. Action is constitutively required for perception. However, importantly, the type of action in question can be either agential or environmental. The problem with behaviourist accounts of perception, such as the one provided by Ryle, was not that they took action to be constitutive of perception. Rather, it was that they took action to be constitutive of perception whilst entirely ignoring environmental activity and focusing solely on agential behaviours.

In short, by accepting a dynamic view of perception which allows that both agential and environmental actions can play a constitutive role in experience, one can accept Ryle’s dissolution of the ‘hard problem of perceptual consciousness’ whilst avoiding the problem of verificationism. Having explained how this argument proceeds, I am now going to explain why RSE is a dynamic view which provides an account of the constitutive role of action in perception that includes both agential and environmental activity. Consequently, I will

conclude that RSE can avoid the ‘hard problem of perceptual consciousness’ whilst providing a phenomenologically convincing account of perceptual experience.

### **§3.2.1: RSE is a Phenomenologically Plausible Dynamical View**

RSE distinguishes between two separate categories of sensorimotor knowledge [Hurley & Noë, 2003; Noë, 2004]:

1. Movement-dependence— this category concerns the law-like relation between the movements of an *organism* and the sensory stimulation it receives from a given object.
2. Object-dependence— this category concerns the law-like relation between the movements of a given *object* and the sensory stimulation received by the organism.

According to RSE, an organism can only be said to perceive an object when its perceptual relation to that object fulfils both the movement-dependent and object-dependent clauses.<sup>28</sup>

Consider my perceptual relation to the dog in front of me. On RSE, I am perceptually related to the dog in front of me because my relation fulfils both the movement- and object-dependent categories of sensorimotor knowledge.<sup>29</sup> My movements will affect my sensory relation to the dog and its movements will affect its sensory relation to me. For example, the dog will loom in my visual field if I move toward it or if it moves toward me. A result of this law-like relation between movement and sensory stimulation is that, were I to leave the

---

<sup>28</sup>For a nice discussion of these two categories of sensorimotor knowledge, which focuses on the (relative) lack of emphasis which has been given to object-dependent sensorimotor knowledge, the interested reader can consult [Cavedon-Taylor, 2011].

<sup>29</sup>I must, of course, also understand this to be the case because I must possess and exercise *knowledge* of this law-like relation between sensation and movement in order to perceive. The following discussion will proceed on the assumption that sensorimotor knowledge is present and being exercised.

café, I would no longer be perceptually related to the dog. Provided we stay in our requisite places (me at home, the dog in the café), there will be no law-like relation between dog sensory stimulation and movement. Therefore, there will be no instantiation of a perceptual relation.

These two categories can not only determine whether an organism is perceptually related to an object. They can also determine the nature of that perceptual relation, and they can do so both inter- and intra- modally [Hurley & Noë, 2003]. The nature of a given sensorimotor relation will determine the difference between modalities of experience (inter-modal differences) and the difference between experiences within a modality (intra-modal differences).

Consider inter-modal differences. Each modality obeys a different set of sensorimotor contingencies because the law-like relation between sensation and movement will differ between modalities. Visually perceived objects will loom as one gets closer to them, they will appear smaller as one moves further away, and they will disappear if one closes one's eyes. Audition does not follow this same pattern of sensorimotor contingencies. If one gets closer to an auditory object it will sound louder, if one moves further away it will sound fainter, and closing one's eyes will have a negligible effect on one's auditory experience. Consequently, we can differentiate modalities by paying attention to the nature of the relevant sensorimotor contingencies.

Consider now intra-modal differences. My visual experience of a rectangular computer screen can be differentiated from my visual experience of a circular coin because my visual relation to each object involves different sensorimotor contingencies. Both objects will loom in my visual field as I move closer toward them. However, the exact nature of the

looming experience differs between each object. The rectangular screen appears trapezoid from a distance, but it appears rectangular as I get closer to it. The coin, however, looks elliptical from a distance but comes to appear circular as it looms in my visual field.<sup>30</sup> Consequently, different visual objects follow different patterns of sensorimotor contingency, and so sensorimotor contingencies can determine intra-modal differences as well.

SE is generally argued for on the basis of its phenomenological plausibility. Sensorimotor contingencies are posited because they help explain perception in a phenomenologically intuitive manner, and this phenomenological intuitiveness is supposed to provide one with a strong reason to accept SE. Susan Hurley and Alva Noë make this point quite explicitly:

When it is brought to our attention that certain sensorimotor contingencies are characteristic of vision, others of hearing, others of touch, there is an ‘aha!’ response. [Hurley & Noë, 2003, p. 146; *cf.* Noë, 2004; Noë, 2012; Ward, 2012; 2015]

We can therefore see that RSE presents one with a phenomenologically plausible account of perceptual experience. Importantly, it provides one with such an account largely because it emphasises the dynamic aspects of perception (the ‘Escher Spaghetti’ relation between action and perception). However, in order to provide a phenomenologically plausible account of perceptual experience which can dissolve the ‘hard problem’, the account being offered must accept Ryle’s view that action is constitutive of perception. Although extant SE theories do not give *environmental action* a constitutive role in perceptual experience

---

<sup>30</sup> One need not accept a representational explanation of perception in order to state that an object appears to be trapezoid from one view-point and rectangular at another. I discuss this point in more detail in chapter seven of this thesis.

(emphasising only *understanding* of this environmental action), there is no reason why they could not do so. SE theories already accept a direct realist account which allows for the environment to play a constitutive role in perception. Therefore, the view can easily be modified to accommodate the additional claim that *environmental action* plays a constitutive role in perceptual experience. I propose that SE should be modified in this manner, and so my own RSE theory does take environmental action to play a constitutive role in perception. By taking environmental action to play this constitutive role, one therefore arrives at a phenomenologically plausible view which takes action to be constitutive of perception whilst avoiding the problem of verificationism.

RSE claims that perceptual experience is constituted by action. However, because both agentive and environmental actions are accorded a constitutive role in perceptual experience, it avoids the problem of verificationism. The key aspect of RSE, which differentiates the view from that of Ryle, is that the movement of *environmental objects* is given a constitutive role in perceptual experience. Ryle also argued that direct perception occurs when one possesses understanding of the law-like relation between sensation and movement [1949/2000, ch. 7]. However, he only ever emphasised the agential aspects of action in perception, and did not give environmental action any role in his theory (presumably because he accepted an input-output view of mentality). This neglect of the constitutive role of environmental movement in perception is what leads Ryle to verificationism. By emphasising the role of sensorimotor contingencies and allowing for environmental movement to play a constitutive role in perception, RSE provides one with a phenomenologically plausible view of perception which can accept that action is constitutive of perception whilst avoiding verificationism. RSE can, therefore, accept Ryle's

dissolution of the ‘hard problem of perceptual consciousness’ whilst providing a phenomenologically adequate conception of experience.

#### **§4: Conclusion**

In this chapter I argued that RSE provides a better explanation of conscious perception than its cognitivist competitors, and that it should be preferred on this basis. I began the chapter by distinguishing between the sub-personal, personal, and conscious levels of explanation and noting that the goal of cognitive science is to explain their inter-relation. In section two, I explained that cognitivist accounts of conscious perception attempt to explain the inter-relation between these levels of explanation via a mechanistic reduction which makes indispensable use of representation. I argued that this kind of explanation is problematic because it gives rise to the ‘hard problem of perceptual consciousness’, and I applied this argument to the case of PP. Finally, in section three, I argued that RSE can provide a phenomenologically plausible account of consciousness which does not give rise to the ‘hard problem of perceptual consciousness’.<sup>31</sup> Cognitivist accounts like PP provide metaphysically problematic accounts of the relation between physical perceptual states and consciousness. RSE provides a phenomenologically plausible account of the relation between physical perceptual states and consciousness which is not metaphysically problematic. It can do so because it adheres to the tenets of *activity* and *knowledge-how*

---

<sup>31</sup>Note that I have not, as yet, provided a positive account as to how RSE proposes to explain the inter-relation between the sub-personal, personal, and conscious levels of explanation. I provide such an account in [ch. 5, §4.2, this thesis]. For the purposes of this chapter, it suffices that I have shown RSE can explain this inter-relation whilst avoiding the ‘hard problem of perceptual consciousness’ whereas cognitivist accounts cannot.



(whereas cognitivist accounts do not). Thus, I conclude that avoidance of this 'hard problem' gives us strong reason to prefer RSE over rival cognitivist theories of conscious perception.

## Chapter Four: Split-Brain Syndrome

In this chapter I am going to argue that RSE can provide a philosophically satisfactory, and empirically plausible, explanation of split-brain syndrome. In part one, I outline the four possible cognitivist interpretations of split-brain syndrome and argue that each is unsatisfactory. I explain that cognitivists are forced into accepting one of these four explanations because they reject the tenets of *activity* and *knowledge-how*. In part two, I argue that RSE can succeed in explaining split-brain syndrome, and that it can do so because it accepts these two tenets. Consequently, I conclude that RSE should be preferred over cognitivist accounts because it provides a parsimonious explanation of split-brain syndrome where cognitivist theories necessarily fail.

## Part One<sup>32</sup>

In part one of this chapter I will explain why cognitivist accounts have problems accounting for the existence of split-brain syndrome and argue that their failure to explain this syndrome presents the paradigm with a serious problem. I begin by explaining what split-brain syndrome is, before summarising the four most common cognitivist explanations of it. I argue that all four cognitivist explanations are unsatisfactory and conclude that, in each case, the unsatisfactory explanation is accepted by theorists only because they have an implicit and a priori commitment to cognitivism.

### **§1: What is a split-brain?**

During the mid-20<sup>th</sup> century patients suffering from severe epilepsy underwent the surgical procedure which I will label the “split-brain procedure”.<sup>33</sup> During this procedure the corpus callosum, which is a band of nerve fibres that connects the brain’s left and right hemispheres, is severed. Severing the corpus callosum helps in the treatment of epilepsy because seizure patterns can no longer spread throughout the brain. Aside from treating epilepsy, the split-brain procedure has the unintended side-effect of removing the main channel through which the brain’s hemispheres communicate with one another.<sup>34</sup> When the

---

<sup>32</sup> Part one of this chapter is largely based on Thomas Nagel’s 1971 essay ‘Brain Bisection and the Unity of Consciousness’. Although written almost fifty years ago, this essay is still one of the best ever written on the philosophical aspects of split-brain syndrome. Throughout part one I have stayed largely true to Nagel’s original analysis of split-brain syndrome, with the only real difference between my summary and Nagel’s being that I discuss contemporary research on the topic where appropriate.

<sup>33</sup> Patients actually underwent one of two procedures— *commissurotomy* (wherein the corpus callosum and some commissures were severed) or *callosotomy* (wherein only the corpus callosum is severed). In what follows, I will use the term “split-brain procedure” to refer to both.

<sup>34</sup> I use the word “communicate” in a non-semantic sense. I will explain in detail in chapters five and six of this thesis how such psychological predicates can be assigned to the brain without thereby committing one to acceptance of sub-personal representation.

corpus callosum is severed the left and right hemispheres become anatomically isolated from one another.<sup>35</sup>

In everyday life split-brain patients are (generally) indistinguishable from their healthy human counter-parts.<sup>36</sup> They claim to feel no different after undergoing surgery, and their friends and families do not notice any differences between the patients pre- and post- surgery. Similarly, animals which have been subjected to split-brain surgery do not appear to behave any differently after it, and their con-specifics do not appear to notice any differences. A particularly interesting example, in this regard, is of a monkey which had undergone the procedure. This monkey happened to be the leader of its troop and, when it was re-introduced into its cage after undergoing surgery, it remained accepted as the leader of its troop [Gazzaniga, 2015]. Thus, split-brain surgery appears to have a negligible effect on the everyday life and behaviour of organisms which have undergone the procedure.

Split-brain procedures were first performed on humans during the 1940s and it was originally thought that the operation had no side-effects. In the 1950s scientists, such as Roger Sperry and Ronald Myers, carried out experiments on animals which had undergone the procedure. These experiments seemed to show that each of the animal's hemispheres was capable of autonomously controlling goal-directed intentional behaviour. The results eventually led scientists to re-test split-brain humans, and in the 1960s it was discovered

---

<sup>35</sup>Sub-cortical connections are unaffected by the procedure [Gazzaniga, 2000].

<sup>36</sup> Strictly speaking this is not exactly accurate. Split-brain patients have exhibited some problems post-surgery, especially regarding cognitive aspects of their lives [Ferguson *et al.* 1985]. However, where perception is concerned, split-brain patients are indistinguishable from both their pre-operative selves and the general population at large. Split-brain patients do sometimes exhibit perceptual difficulties. However, these difficulties are caused by pre-existing neuropsychological illnesses which preceded the surgery.

that these subjects' hemispheres could also autonomously control goal-directed behaviour during experimental conditions.

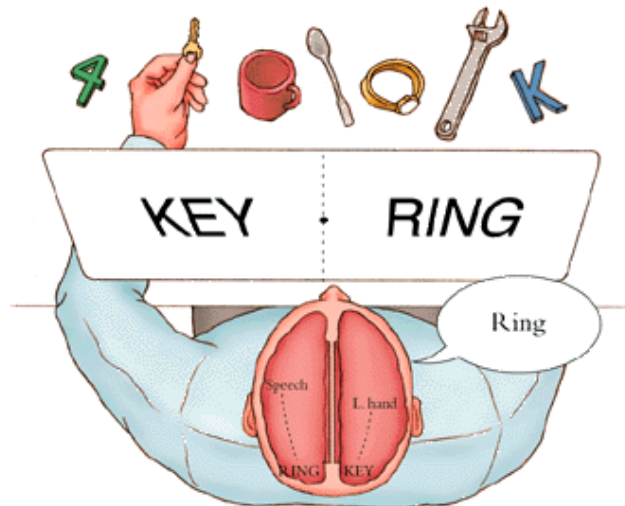
### ***§1.1: What happens in a split-brain experiment?***

During a typical split-brain experiment, the subject will be seated in front of a screen and instructed to fixate on a central marker. By having the subject fixate on a central marker the experimenter can ensure that subsequently presented stimuli will be localised to a single hemisphere, and so the information provided by the stimuli will be available to only one of the hemispheres. The subject is then asked to engage in a goal-directed behaviour which is linked to the stimulus. For example, they will be asked to provide a verbal report of what they perceived. Provided the experiment has been carried out correctly and the experimental stimuli were localised to a single hemisphere, the subject will engage in autonomous goal-directed behaviours, with one of these behaviours being controlled by the left hemisphere and the other controlled by the right.<sup>37</sup>

---

<sup>37</sup> In what follows I will focus on visual split-brain experiments. However, split-brain experiments have also been successfully conducted in the modalities of audition, smell, and (to a lesser extent) touch.

Figure Five: Sample Split-Brain Experiment



[Bayne, 2010, Figure 9.1]

In one representative experiment the word “Key-Ring” is flashed onto a screen. Because the subject is fixating on a central point, the left hemisphere is presented with visual stimulus only of the word “Ring”, whilst the right hemisphere is presented with visual stimulus only of the word “Key”. The subject is then asked to verbally report the word they perceive and to use their left hand to pick up the object which best corresponds to the word perceived. The patient will verbally report that they have seen the word “Ring”. However, they will pick up a key with their left hand. When the same experiment is carried out on healthy subjects this kind of dissociation will not be found. Subjects will simply report that they have seen the word “Key-Ring”, and select a key-ring from the set of available objects.

Split-brain patients behave in this strange manner because each of their hemispheres is informationally isolated from the other. Neither hemisphere has access to the visual stimulus processed by the other, and so when the subject is asked to engage in goal-directed behaviour in response to the stimulus perceived each hemisphere supports intentional action related to the stimulus it has been exposed to.

There are five main interpretations of split-brain syndrome:

1. *Interpreter Model*: consciousness is confined to the language-using left hemisphere
2. *Two-Streams Model*: each hemisphere supports a separate stream of consciousness. The contents of each stream may be duplicated.
3. *Partial Unity Model*: the split-brain patient possesses one partially unified stream of consciousness.
4. *Switch Model*: there is one stream of consciousness and the supervenience base of this stream switches between hemispheres.
5. *Experimental Aberration*: the experimental condition itself causes a split in consciousness.

The focus of part one of this chapter will be on interpretations one to four. I will argue against each of these interpretations and will explain why theorists accept these interpretations, in spite of their deficiencies, largely because they have an a priori commitment to cognitivism. I will therefore conclude that cognitivism is incapable of providing a satisfactory explanation of split-brain syndrome.

## **§2: The Interpreter Model**

Proponents of the *interpreter model* argue that the split-brain experiments reveal a pre-existing fact— consciousness is confined to the language-using left hemisphere [Dennett, 1993; Eccles, 1965; Gazzaniga & Le Doux, 1978; Gazzaniga, 2000]. The interpreter model can be helpfully understood as proposing a thesis similar to one which can be found in Norman Malcolm's *Dreaming* [1959]. Malcolm argued that what we call dreams are actually a set of just-so stories that we tell ourselves, about experiences we think we had, when we were asleep. According to the interpreter model, consciousness is constituted by a set of stories about experiences which we think we had that are strung together on the basis of our behaviour. The model is primarily motivated by the idea that the actions which the right hemisphere engages in do not require consciousness (because simple computers or zombie mechanisms in the brain could perform similar behaviours, and we do not consider these to be conscious).<sup>38</sup> Therefore, the distinctively human aspects of split-brain behaviour (which are best explained as the product of consciousness) are thought to be those which originate from the language-using left hemisphere.

Consider a set of experiments carried out on split-brain patient J.W. [Gazzaniga, 2000, fig. 19]. J.W.'s right hemisphere was exposed to a picture of a chicken claw and his left hemisphere was exposed to a picture of a snow scene. When asked to pick from a set of picture cards the one which fitted most closely with the stimulus he had perceived, J.W. selected a picture of a chicken with his left hand and a picture of a shovel with his right. In a different experiment, J.W.'s left hemisphere was exposed to the word "Music" and his right

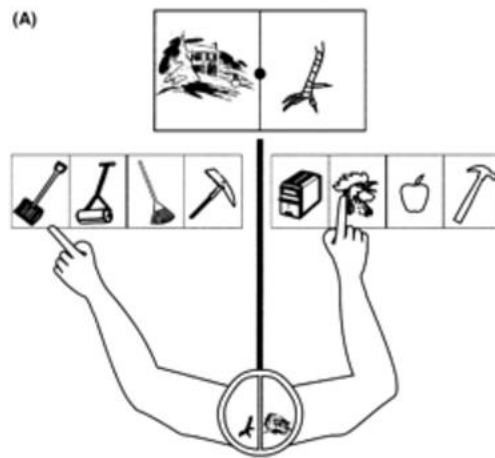
---

<sup>38</sup> The interpreter model originally gained prominence through the work of Sir John Eccles, and was motivated in part because theorists had been unable to elicit responses from the 'dumb' right hemisphere during early split-brain research. However, when experimental techniques improved it became clear that the right hemisphere can support sophisticated cognitive behaviour.



hemisphere was exposed to the word “Bell”. When asked to select the picture card which best corresponded to the stimulus perceived with his left hand, J.W. picked out a picture of a bell.

Figure Six: Illustration of experiment performed on J.W.



[Gazzaniga, 2000, fig. 19]

The best explanation of J.W.’s behaviour in the first experiment is that both of his hemispheres were acting autonomously— J.W. picked out a chicken with his right hand because his left hemisphere had been exposed to a chicken claw, and he picked out a shovel with his left hand because his right hemisphere had been exposed to a picture of snow. However, when asked to explain why he had selected the two cards that he did, J.W. provided a surprising explanation— he (correctly) attributed his selection of a chicken card to his having perceived a chicken claw, but he (incorrectly) attributed his selection of a shovel to the idea that “you need a shovel to clean out the chicken shed” [Gazzaniga, 2000,

p. 1318]. The best explanation of the results of the second experiment is that J.W. picked a picture of a bell with his left hand because his right hemisphere had been exposed to the word “Bell”. However, when asked to provide his reasons for selecting a bell card, J.W. answered as follows: “Music—last time I heard any music was from the bells outside here, banging away” [Gazzaniga, 2000, p. 1318]. J.W.’s language-using left hemisphere is aware that he has been exposed to the word “Music” but it is not aware his right hemisphere has been exposed to the word “Bell”. It therefore explains his selection of a bell card with a false rationalisation— that his left hand picked out a bell because he can often hear bells ringing when he sits in the room where the experiments are conducted.

In both of these cases J.W. has provided a false explanation of his behaviour. The clinical name for such post-hoc rationalisation of behaviour is “*confabulation*”. The proponents of the interpreter model believe that this kind of confabulation is in fact a pervasive aspect of human activity. They believe that we engage in this kind of behaviour *all the time*, and that its exhibition is not restricted solely to neuropsychological maladies and forced experimental set-ups. A good example of why theorists think we should believe confabulation to be a pervasive aspect of the human condition can be found in the experimental paradigm of *choice blindness* [Johansson *et al.*, 2005; 2006; 2008].<sup>39</sup>

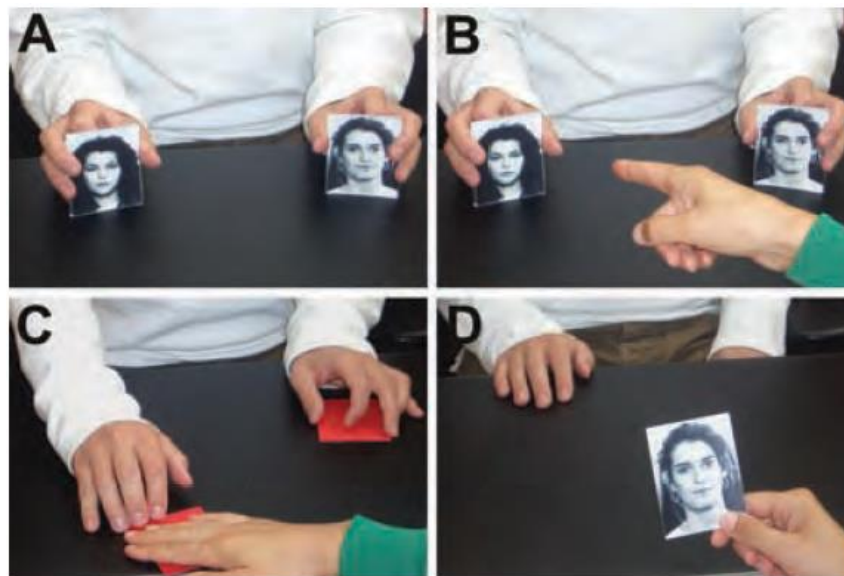
In a representative experiment (see figure seven) subjects were shown pictures of two female faces (A) and asked to pick which face they found more attractive (B). Once a card had been picked the experimenter handed it over to the subject, and the subject was asked to explain why they had chosen the face that they did (C and D). Unbeknownst to the

---

<sup>39</sup> To the best of my knowledge no proponents of the interpreter model use evidence from choice blindness to argue for their claim (preferring instead other examples from the psychological literature). I prefer to use this example because I think it provides a particularly compelling argument for the confabulation thesis, and it is this thesis which ultimately motivates and underlies the interpreter model.

subject, however, the two cards had been surreptitiously switched by the experimenter during the handover. As such, the subject was handed the card which they had *not* selected. Subjects proceeded to explain why they had chosen as they did, and were completely oblivious to the fact that they had not chosen the card they were currently holding. In short, subjects in this experiment provide reasons for having chosen a card which they had not actually chosen— they are confabulating.

Figure Seven: Choice Blindness Experiment



[Johansson *et al.* 2005, fig. 1]

The same experiment was run on a control group, but this time the experimenter did not switch the card chosen during handover. When the experimenters compared the explanations provided by the subjects in both versions of the experiment they could find no

dissimilarities between the two. On the basis of these results Johansson *et al.* concluded as follows:

Thus, the lack of differentiation between the manipulated and non-manipulated reports could be seen to cast doubt on the origin of the non-manipulated reports as well; confabulation could be seen to be the norm and truthful reporting something that needs to be argued for. [Johansson, 2008, pp. 1-2]

These results support the interpreter model. They appear to show that even healthy humans engage in post-hoc rationalisations of their behaviour. Such choice blindness experiments are representative of the kind of empirical work which is taken to provide evidence for the interpreter model.

The two most prominent proponents of the interpreter model, Daniel Dennett and Michael Gazzaniga, believe there to be an evolutionary explanation for the existence of our 'left hemisphere interpreter'. Each of these theorists takes mentality to be constituted by an anarchic assortment of de-centralised, disconnected, and distributed cognitive processes [Dennett, 1993; 2005; Gazzaniga, 1978; 2000; 2012]. As a consequence, because there is no centralised planner or agent in control of our behaviour, the behaviour we engage in is determined rather arbitrarily. However, although this is true, it does not feel to us as if our behaviour is arbitrary. It feels to us as if there is a persisting 'self' or 'I' which is responsible for our actions. The interpreter model is invoked to explain this (illusory) feeling, and explains it to be based upon our language-using left hemisphere. The left hemisphere is thought to create a running story that pieces our arbitrary and disparate actions into some semblance of a coherent and goal-directed narrative, and it thus provides us with the illusion of a persisting 'self' or 'I'. Dennett and Gazzaniga are led to this conclusion because

they accept a cognitivist view of the mind. Each of these theorists believe that the mind is constituted by brain-based representations. The empirical evidence appears to show that our brain is extremely fragmented. Therefore, it is concluded by both theorists that we ourselves are extremely fragmented (because we are our brain, and our brain is fragmented).<sup>40</sup>

To sum up— the interpreter model takes split-brain syndrome to unveil a pre-existing aspect of conscious experience. Split-brain experiments show that human beings do sometimes provide post-hoc and confabulatory rationalisations of their behaviour. However, there is empirical evidence that such post-hoc confabulation is present even in healthy human subjects. These post-hoc rationalisations are thought to be created by the language-using left hemisphere, and it is believed that the resulting verbal reports are constitutive of conscious experience. Therefore, split-brain experiments simply unmask a pre-existing fact about human experience— it is localised to the language-using left hemisphere.<sup>41</sup>

---

<sup>40</sup> It may be objected that these theorists can't be arguing in such a fallacious manner. By the same reasoning we can come to the conclusion that we weigh around 10 grams:

1. My brain weighs ten grams
2. I am my brain
- Therefore
3. I weigh ten grams.

It would be especially surprising if Dennett (who is famous for arguing that there is a distinction between vehicles and content, and for arguing that we do not need to posit an isomorphism between vehicle and content) was arguing in this manner. However, it is the case that Dennett and Gazzaniga are arguing in exactly this manner— presumably because they are cognitivists (*cf.* Thompson [2007], who argues that Dennett is guilty of exactly this kind of reasoning in his work on the mental imagery debate).

<sup>41</sup> The reader may be wondering whether I have confused Dennett and Gazzaniga's discussion of consciousness and their discussion of the self— I have sometimes been talking of one, and sometimes talking of the other, as if the two are synonymous (of course, they are not). This ambiguity is present in the work of Dennett and Gazzaniga themselves and so, since I am simply outlining their views here, I have retained the ambiguity present in their own writings.

### §2.1: Problems with Interpreter Model

The main problem with the interpreter model is that it is phenomenologically implausible. It is not at all clear how one could connect the idea that consciousness is constituted by narratives created by the language-using left hemisphere with our phenomenological experience. Indeed, Dennett's attempt to flesh this idea out results in the conclusion that we are wholly mistaken about the nature of our phenomenological experience:

Exactly! *There seems to be phenomenology*. That's a fact that the heterophenomenologist enthusiastically concedes. But it does *not* follow from this undeniable, universally attested fact that *there really* is phenomenology. This is the crux. [Dennett, 1993, p. 366, *italics in original*]

I *am* denying that there are any such properties. But (here comes that theme again) I agree wholeheartedly that there seem to be qualia. [Dennett, 1993, p. 372, *italics in original*].

Proponents of the interpreter model arrive at the conclusion that we are wholly misguided about the nature of our conscious experience because they assert that consciousness is constituted by linguistic report. However, I don't think there are any good reasons to accept this assertion. For one, accepting it requires denying conscious experience to any organism which does not possess the ability to use language. This would require denying that all non-human animals, and most human infants under the age of four years old, are conscious. Such a conclusion is at least *prima-facie* implausible.<sup>42</sup>

The ability to use language *may* radically transform the nature of one's phenomenological experience, but it does not follow from this that linguistic report itself

---

<sup>42</sup> Furthermore, it is arguable that attributing consciousness on the basis of intentional agency (as opposed to linguistic report) can account for all of the empirical evidence amassed in support of the interpreter model. I do not pursue this point here, but the interested reader can consult [Bayne, 2010, pp. 98-101; 2013] for an argument to this effect.

constitutes or is identical to our phenomenological experience. At most, the empirical evidence shows that humans are (at times) extremely poor introspectors [Bayne, 2010, ch. 4, §4.1-4.3]. Provided we reject the assumption that consciousness is constituted by linguistic report, the existence of confabulation has no direct implications for the existence or nature of conscious experience itself. Rather, it has implications only for the nature of our reports about that experience. Therefore, we can see that the empirical evidence amassed in support of the interpreter model does not require acceptance of this model.

In short, we have two strong reasons to reject the interpreter model: (1) it requires that only linguistic organisms are considered conscious; (2) it requires an extremely implausible, phenomenologically revisionary conception of experience. Furthermore, the evidence amassed in support of this thesis does not uniquely specify, or require acceptance of, the interpreter model. I therefore conclude that the interpreter model explanation of split-brain experiments should be rejected.

A deeper, more important point to note here is that the interpreter model is advanced largely because its proponents are working within a cognitivist framework. Proponents of this model believe that the conscious mind is constituted by brain-based representational vehicles. As such, the primary motivation of their theorising is not simply to provide the best explanation of the conscious experience of split-brain patients and healthy humans. Rather, it is to provide the best explanation of these subjects' conscious experience which is compatible with a cognitivist view of the mind. Thus, these theorist's choice of possible explanations for split-brain syndrome is restricted to those which allow for consciousness to occur solely 'in the head'.

### **§3: Two-streams Model**

According to the *two-streams* model of split-brain syndrome, by splitting the brain during surgery the surgeon creates a split in consciousness. Patients are thought to possess one stream of consciousness before surgery, and to possess two streams of consciousness after it [Sperry 1966; 1968; Parfit, 1984]. According to Roland Puccetti's [1981] version of the two-streams model, the split-brain procedure exposes a pre-existing condition— each hemisphere of the human brain has always supported its own stream of conscious experience, and the split-brain operation simply makes this fact salient.<sup>43</sup>

Proponents of this model argue that split-brain subjects do not display aberrant behaviour in everyday circumstances, in spite of the fact that they now support two separate streams of consciousness, because the contents of both streams is identical. Both hemispheres are housed in the same body, both encounter the same environments, both share the same blood-stream, and both share many of the same (sub-cortical) connections. Consequently, it is argued that each hemisphere will support separate streams of consciousness with identical contents, and so subjects only engage in aberrant behaviour during split-brain experiments because these experiments serve to temporarily cause each stream to diverge in content from the other.

---

<sup>43</sup> Interestingly, Roger Sperry does sometimes appear to say something quite similar to Puccetti. For example— “In a normal person, also, it would appear that there must be much the same sort of double sensing by the left and also by the right hemisphere- at least there is no way to rule it out.” [Sperry, 1966, p. 303]. However, he is commonly interpreted as accepting a model upon which the split of consciousness occurs only after the split-brain procedure, and most of the comments in his writings support this interpretation (for example, almost immediately after writing the preceding Sperry says: “The conscious awareness of the minor hemisphere *produced* by this vertical splitting.” [*ibid.* p. 303, *my emphasis*]).



### ***§3.1: Problems for the Two-streams model***

Jerry Fodor famously remarked that:

[I]f commonsense psychology were to collapse, that would be, beyond comparison, the greatest intellectual catastrophe in the history of our species. [Fodor, 1987, xii]

I demur. In my opinion, the negative consequences of the (potential) falsity of folk psychology pale in comparison to the negative consequences which would ensue were the two-streams model to turn out correct. I submit *that* “would be, beyond comparison, the greatest intellectual catastrophe in the history of our species”. My reason for thinking this is simple— if this model were correct then *everything* we have thought about consciousness up until the discovery of the effects of split-brain surgery would be shown to be false. It does not seem to split-brain subjects as if they house two separate spheres of consciousness, nor does it seem to healthy subjects that they possess two separate spheres of consciousness. Moreover, before the discovery of split-brain syndrome we had *absolutely no reason* to think that we could house or support two separate streams of consciousness. Fortunately for our self-conception, however, I don’t believe that the evidence supporting the two-streams model is as strong as it first appears.

The biggest problem with the two-streams model is that it has difficulty accounting for the fact that, bar specialised experimental circumstances and isolated incidents outside the lab, split-brain subjects exhibit behaviour which is indistinguishable both from their pre-surgery selves and indeed the general population at large [Bayne, 2010, ch. 9; Nagel, 1971; Tye, 2003, ch. 5]. The fact that it took scientists *twenty years* to discover the bizarre effects

of split-brain surgery, in spite of having tested patients post-surgery, is itself a datum which requires explanation. Tim Bayne provides a succinct summary of this problem:

Given that the two hemispheres have different memory stores and cognitive styles, even complete mirroring of content might be expected to lead to behavioural disunity were both hemispheres conscious in parallel. [Bayne, 2010, p. 205]

It is a well-documented fact that human brains are lateralised for many functions [Gazzaniga & Le Doux, 1978, ch. 2; Gazzaniga, 2000]. Each hemisphere possesses the ability to support a variety of functions which are not present in, or supported by, the opposite hemisphere.<sup>44</sup> For example, the main areas associated with language (such as Broca's area, responsible for speech production, and Wernicke's area, responsible for speech comprehension) are located in the left hemisphere in 90-95% of right handed human beings, and in 70% of left handed and ambidextrous human beings [Rasmussen & Milner, 1977]. After a certain developmental point in time, destruction of these areas leads to the loss of the ability in question (for example, lesion of Broca's area results in Broca's aphasia, wherein the patient loses the ability to speak).

The two-streams model explains the cohesive everyday behaviour of split-brain subjects as based on a mirroring of content in the two separately conscious streams. However, given that the human brain is lateralised for a number of functions (such as

---

<sup>44</sup> Though we should be wary of taking this point too far. It is now a somewhat pervasive myth in our culture, and in some 'pop' science, that we possess two separate styles of brain — an analytic left-brain and an emotional and creative right-brain. As argued by Gazzaniga and Le Doux [1978, ch. 2], the experimental data do not back up this view. Indeed, many of the experiments which gave birth to this myth could be explained purely on the basis of the means by which split-brain subjects had to respond to the tasks (with their hands). Once this response problem had been resolved, it turned out that the left hemisphere possesses only a slight advantage over the right in analytic tasks, whilst the right possesses only a slight advantage over the left in tasks involving creative reasoning.

language) it is difficult to understand how each stream could possess identical contents.<sup>45</sup> For example, the conscious stream supported by the left hemisphere will possess linguistic abilities whereas the conscious stream supported by the right hemisphere will not. Given that the possession of linguistic abilities serves to radically bootstrap an organism's cognitive abilities [Clark, 1997; 2008], we should expect at least *some* behavioural disunity to be exhibited by the split-brain subject, at least *some* of the time, during its everyday activities. However, this kind of disunified behaviour simply is not exhibited by split-brain subjects frequently enough to provide compelling evidence for the two-streams hypothesis.<sup>46</sup>

No proponent of the two-streams model has ever provided a satisfactory response to this objection. Having made this argument, however, I do not take myself to have provided a refutation of the two-streams model. Rather, I take myself to have made the following *weaker* argument—the two-streams model can make sense of an organism's behaviour *during* split-brain experiments, but it cannot make sense of that same organism's behaviour outside experimental contexts. As such, this model provides an unsatisfactory explanation of split-brain syndrome.

A deeper point worth noting here is that, once again, proponents of the two-streams model argue for their view largely because they accept a cognitivist framework. Proponents of this model assume that the brain is constitutive of consciousness because the vehicles of

---

<sup>45</sup> Note, this point applies even if the difference between the two hemispheres has been exaggerated. For example, even if the *only* difference between the two hemispheres was the presence of mechanisms which sub-serve linguistic functions, we would still expect a notable divergence of behaviour at least occasionally.

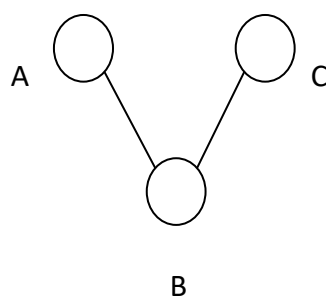
<sup>46</sup> Split-brain patients do sometimes engage in aberrant behaviour outside of experimental contexts. For example, one hand may attempt to button up a shirt whilst the other hand attempts to unbutton it. However, these behaviours are extremely rare, and they tend to disappear entirely a couple of months after surgery. I discuss this point in more detail in part two of this chapter.

conscious representational states are housed there. Given that the brain is considered to be constitutive of consciousness, when the brain is split it follows that consciousness will also split. Once more, an a priori assumption of cognitivism is constraining the kinds of explanations theorists think they can provide. The two-streams model is offered up to explain split-brain syndrome not because it provides the most satisfactory explanation of the syndrome. Rather, it is offered up because theorists take it to provide the best explanation of split-brain syndrome which is compatible with cognitivism.

#### **§4: Partial Unity**

The *partial unity* model can be seen as a milder version of the two-streams model, and as one that avoids many of the two-stream model's excesses.<sup>47</sup> On the partial unity interpretation, consciousness is conceptualised as possessing a branching structure:

**Figure Eight:** Partial Unity Model



[Schechter, 2014, fig. 2]

<sup>47</sup> Surprisingly this model has had few proponents. With the exception of Nagel [1971] (who only gestures at the model), Michael Lockwood (who accepts the model [1989] whilst also admitting that he has serious doubts about whether he truly believes it [1994]), and Elizabeth Schechter [2014], I am not aware of any other proponents of partial unity.

For example, imagine that a split-brain subject is presented with stimulus *A* to their left hemisphere (the word “Ring”) and stimulus *C* to their right hemisphere (the word “Key”). In addition, imagine that stimulus *B* has been presented to both hemispheres (for example, *B* may be a tactile sensation experienced on the subject’s neck).<sup>48</sup> If the subject’s conscious state is partially unified then *A* and *C* will be experienced together with *B*, but *A* and *C* will not be experienced together with each other. Consequently, although the subject does only harbour one stream of conscious experience, the contents of this stream are only partially unified with one another.

Michael Lockwood [1989, p. 87-94] is generally credited as the first theorist to propose the partial unity model.<sup>49</sup> He motivated his model by referencing a problem with the two-streams model— it presents us with an unrealistic binary option. Either consciousness is split, or it is not split. The two-streams model requires that there is a specific occasion upon which the split in conscious stream occurs. Lockwood noted, however, that any occasion selected by a two-streams theorist as the moment in which consciousness splits will be entirely arbitrary. For example, does the split occur immediately following the severing of one fibre? Or does it occur after more than half of the corpus callosum has been severed? Lockwood argued that any answer to this question would be wholly arbitrary. As such, the motivation behind the partial unity model is that conscious unity is not an all-or-nothing affair and that accepting partial unity provides a more realistic and parsimonious explanation of split-brain syndrome. Furthermore, Lockwood argued that his partial unity model seems particularly plausible when one notes that much of the

---

<sup>48</sup> The neck maintains connections with both hemispheres even after split-brain surgery.

<sup>49</sup> He does note that similar views were present in the neuropsychological literature around the time he was proposing his model. However, Lockwood was unaware of these views at the time, and in any case, it is not clear that the partial unity model was being *explicitly* advocated in this literature [Lockwood, 1994, p. 72].

subject's brain remains intact and connected even after surgery (for example, the brain-stem and sub-cortical structures remain connected to both hemispheres, both hemispheres share the same blood supply, both hemispheres' sleep-wake cycles converge, and so on).

#### ***§4.1: Problems for Partial Unity***

The partial unity model has traditionally been attacked from one of two angles:

1. Partial unity itself is impossible to conceive of.
2. There is no substantive difference between partially unified and duplicate streams of consciousness. Any appearance of a difference can be explained as merely verbal.

Objection (1) is motivated by the idea that, if there is only one stream of consciousness, then it is conceptually incoherent to talk about partial unity. The very idea of partial unity is nonsensical because it is impossible to conceive of what it would be like to possess a partially unified conscious field. For example, how could someone possess an experience of *A*, *B*, and *C*, with *A* and *C* being consciously experienced alongside *B* but not each other?

What would such an experience feel like? According to proponents of this objection, the very idea of such an experience is incoherent [Bayne, 2010; Dainton, 2000; Nagel, 1971].

This objection is commonly levelled at proponents of the partial unity model. However, I do not think it is convincing, because the subjective point of view taken alone is not capable of deciding this matter [Hurley, 1998, ch. 5; 2003].

If we concern ourselves solely with a subjective point of view, then there is absolutely no difference (phenomenologically speaking) between fully unified, partially

unified, or split streams of consciousness. Restricting ourselves to the first-person perspective provides us with absolutely no way of differentiating between these three states— one cannot tell whether one possess one unified, one partially unified, or two duplicate streams of consciousness, solely on the basis of phenomenological reflection. Susan Hurley argues that this conclusion requires us to look toward objective evidence in order to determine whether or not the partial unity model is correct.

She argues that when we do look toward objective evidence we will still be unable to differentiate between the partial unity interpretation and rival models. According to Hurley, there is only one objective factor which could help differentiate between these models— neural structure. If the brain's structure is considered isomorphic to the structure of one's conscious field, then one will be able to tell whether one's conscious field is unified, partially unified, or duplicated. A fully unified neural structure indicates a fully unified conscious field, a partially unified structure indicates a partially unified conscious field, and so on. Hurley rejects this neural isomorphism constraint using her *hypothetical acallosal* thought experiment. This thought experiment (which will be explained in more detail in part two of this chapter) appears to show that it is possible for the same states of conscious unity to be supported by entirely different neural structures. Therefore, neural isomorphism cannot be used to determine the truth or falsity of the partial unity model.

Elizabeth Schechter responds to Hurley's argument by noting that it has just as much force as an argument against the two-streams model as it does as an argument against the partial unity model [2014, §6]. Hurley's point is that we should look toward objective, empirical evidence (neuroanatomical or neurofunctional structure) in order to determine the truth or falsity of the partial unity model. However, the method she advocates does not

allow for this objective, empirical evidence to play any role in determining the truth or falsity of partial unity. Schechter contends that exactly the same argument can be supplied against the two-streams model— if neuroanatomical factors cannot be used to determine the structure of a conscious state, and if subjective factors are silent on the matter, then it follows that neither of these factors can be used to determine the truth or falsity of the two-streams model. Schechter therefore concludes that Hurley’s argument fails because it entails far too strong a conclusion— it requires an a priori rejection of two empirical possibilities. As such, she thinks that we should reject Hurley’s argument.

I agree with Schechter’s assessment of Hurley’s argument, and so I will accept that the partial unity model may turn out to provide the best explanation of split-brain experiments and, indeed, that it may even provide the best explanation of the structure of split-brain subjects’ conscious states during everyday life. Consequently, I will leave it on the table for now as an empirical possibility. In part two of this chapter, I will explain how the insights of the partial unity model can be accepted within an RSE framework without requiring that split-brain consciousness is always partially unified.

### **§5: Switch Model**

According to the *switch model* of conscious unity, each subject possesses a single stream of conscious experience even during split-brain experiments. The switch model takes the supervenience base of this single conscious stream to alternate between hemispheres during split-brain experiments. For example, the subject of the experiment is conscious of the word “Key”, then the word “Ring”, then the word “Key”, and so on, because the supervenience base of their single stream of consciousness is constantly switching between



the two hemispheres. This model's two most prominent exponents are Tim Bayne and Susan Hurley. Bayne argues that an attentional mechanism in the brain is responsible for initiating the switch between hemispheres [2010],<sup>50</sup> whilst Hurley argues that motor intentions are responsible for initiating the switch between hemispheres [1998, ch. 5; 2003]. A major virtue of this account is that it can provide a good explanation as to why split-brain subjects exhibit normal behaviour in everyday life. Even though the supervenience base of their conscious stream is constantly switching between each hemisphere, their mental life is contiguous because they possess a single contiguous stream of conscious experience.<sup>51</sup>

### ***§5.1: Problems with the Switch Model***

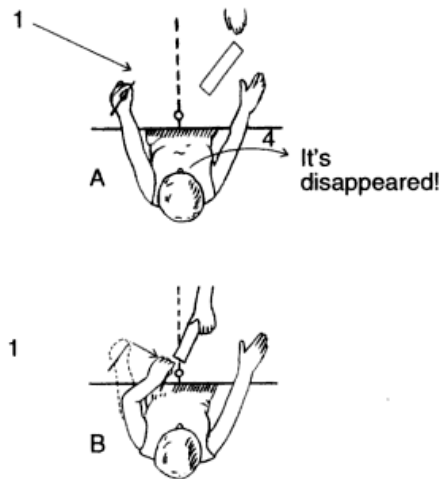
Susan Hurley uses an experiment carried out by Colin Trevarthen [1974] to motivate her version of the switch model [1998, p. 170-190]. In this experiment, the split-brain subject's left hemisphere was exposed to a piece of white paper whilst their right hemisphere was not exposed to any stimulus. They were then instructed to pick up the paper with their left hand. Although the paper was initially visible to the subject, it vanished from view as soon as they attempted to pick it up with their left hand. Presumably, the paper vanished from view because the left hand is controlled by the right hemisphere, and the right hemisphere does not have perceptual access to the piece of paper.

---

<sup>50</sup> Bayne provides mixed messages on this point. He notes that it could be the case that "everyday life may permit patients to enter into a state of non-focal awareness...although focal consciousness is restricted to a single hemisphere at a time, it might be possible for non-focal awareness to be distributed across the patient's two hemispheres." [Bayne, 2010, p. 214]. As such, he leaves open the possibility that the switch model may apply only during experimental contexts, or those which require the subject to focus attention on a given task.

<sup>51</sup> It is worth noting that healthy subjects should not be expected to exhibit this switching behaviour because their corpus callosum has not been severed. As such, the switch model requires that severing the corpus callosum results in a permanent change to the structure of conscious experience.

Figure Nine: Trevarthen's Experiment



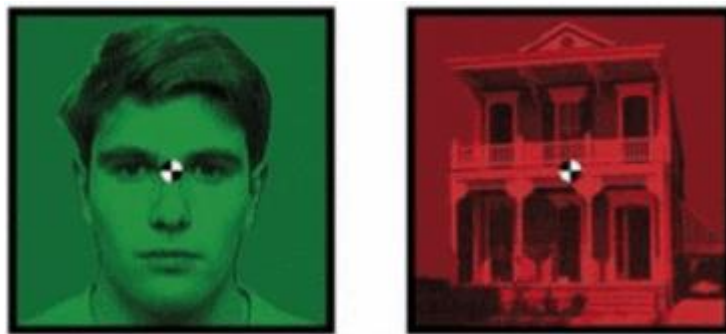
[Trevarthen, 1974, p. 197; referenced in Hurley, 1998, p. 171]

Hurley explains this experiment as follows— the paper vanishes from experience because the subject's motor intention to pick up the paper switches the supervenience base of their conscious stream across hemispheres. The subject can initially report an experience of the paper because it is located in their right visual field and, consequently, information about its presence is available to the language using left hemisphere. However, the motor intention to pick up the paper with the left hand originates from the right hemisphere. This switches the supervenience base of the conscious stream to the right hemisphere, which has no access to the piece of paper, and it is for this reason that the piece of paper disappears from the subject's experience. This pattern (of a piece of paper disappearing when attempts to grab it are made) recurs until the paper is moved by the experimenter into the left visual

field. As soon as the paper is moved into this area, it is immediately snatched up by the waiting left hand.

The switch model can be understood as presenting a view of conscious experience which is quite similar to the phenomenon of binocular rivalry. In binocular rivalry experiments, separate images are projected to a subject's left and right eyes. For example, the left eye is exposed to a picture of a face whilst the right eye is exposed to a picture of a house. Subjects of this experiment report an experience of a face or a house, with the stimulus being perceived constantly switching.<sup>52</sup>

Figure Ten: Binocular Rivalry



[Blake and Tong, 2008, fig. 2]

It should be obvious that, according to the switch model, the experience of split-brain patients is extremely similar to the experience of subjects of binocular rivalry experiments.

---

<sup>52</sup>I discuss binocular rivalry in more detail in chapter five.

In both cases, there is a switch in the contents of the subject's conscious stream. However, there is an important difference between the two types of experience— in binocular rivalry subjects are aware of the switch in their conscious experiences, whereas in split-brain experiments subjects are (usually) not aware of a switch in their conscious experiences. This observation presents the proponent of the switch model with a problem. If the subject's conscious stream is constantly in flux during split-brain experiments (in much the same way it is in flux during binocular rivalry experiments), then why is the switch in experience detected by subjects of rivalry but not detected by subjects of split-brain experiments?

Tim Bayne responds to this problem by referencing the psychological literature on the unreliability of introspection and comparing split-brain patients to people suffering from hemispatial neglect [2010, pp. 218-220]. Hemispatial neglect (which was discussed in chapter two) is an attentional disorder which prevents patients from becoming aware of one half of their visual field (usually the left half). People suffering from this disorder, however, are completely unaware of it. They do not realise that they lack awareness of an entire half of their visual field. Bayne contends that split-brain patients are in an exactly analogous situation. He argues that they are totally unaware of the constant switching of their experiential states (perhaps because they have brain damage), and so takes himself to have explained why split-brain patients do not realise their conscious states are constantly switching between the two hemispheres.

I reject Bayne's response because it is unparsimonious. We have just seen that there are split-brain experiments wherein the subject *was* aware that a switch in conscious experience had occurred. Bayne's explanation could therefore not be correct, because split-brain patients do sometimes realise that a switch in their conscious stream has occurred. It

isn't clear to me how Bayne (or any other proponent of the switch model) could provide a parsimonious explanation of the difference between cases wherein split-brain patients are aware of the switch in their conscious streams, and cases wherein they are not. The switch model should therefore be considered problematic because it cannot account for the experimental data in an even-handed manner.

A deeper point, which has surfaced in the discussion of each account of split-brain syndrome countenanced thus far, is that Bayne accepts the switch model largely because he accepts cognitivism. Bayne does not accept the switch model solely because he thinks it provides the best account of split-brain syndrome.<sup>53</sup> Rather, he accepts the switch model because he takes it to provide the best account of split-brain syndrome which is compatible with cognitivism.<sup>54</sup> Once more, we can see that an a priori acceptance of cognitivism is playing a huge role in the accounts of split-brain syndrome that theorists take to be acceptable.

## **§6: Conclusion of Part One**

In part one of this chapter I surveyed the four most common explanations of split-brain syndrome and argued that each was unsatisfactory. I rejected the interpreter model because it relied on the assumption that consciousness is constituted by linguistic report

---

<sup>53</sup> Bayne advances his switch model in the context of providing an overall account of conscious unity. He therefore does not supply his account solely to explain split-brain syndrome. Rather, he uses the example of split-brain syndrome as a test-case for his account. His overall account of conscious unity is itself, however, primarily motivated by cognitivist principles.

<sup>54</sup> It is worth noting that Hurley also accepts a version of the switch-model and yet rejects traditional cognitivist accounts of consciousness (because she argues for externalism about the representational vehicles of experience [1998]). Her position therefore is an exception to the general pattern observed in part one of this chapter (that theories of split-brain consciousness are accepted, in spite of their problems, largely because they are compatible with cognitivism).

and so is phenomenologically implausible. Although I did not reject the two-streams and partial unity models outright, I did provide reason to think they will be unable to account for the unified behaviour of split-brain subjects during everyday life. Finally, I rejected the switch model because its explanation of split-brain experiments is unparsimonious.

Throughout part one of this chapter a deeper underlying theme was uncovered—each of these views is motivated largely because the people who propose them are working within a cognitivist paradigm. Each theorist believes that the conscious mind is brain-bound, and so each attempts to explain split-brain syndrome in a manner compatible with internalism. They are willing to accept the limitations of their respective views largely because they are in the grip of a cognitivist view of mind, and so can see no better way to account for split-brain syndrome.

Perhaps we should therefore stop artificially restricting our options and look toward other possible explanations of split-brain syndrome. Such explanations go beyond the confines of the skull and look toward the body in which the brain is situated as well as the environment in which it is embedded. In part two, I will propose one such account and argue that my own theory of perceptual consciousness (RSE) can provide a parsimonious explanation of split-brain syndrome. This argument is largely empirical in nature, and necessarily speculative (because the relevant empirical work has simply not been carried out). That noted, I do offer predictions which could be tested empirically and my argument does rely on empirical evidence. I will conclude that my alternative explanation provides us with positive reasons to take accounts of consciousness predicated on *activity* and *knowledge-how* seriously as empirical hypotheses— they show promise of being able to

account for the existence of conscious experiences which cognitivist accounts are incapable of explaining.

## Part Two: RSEcuing Split-Brain Syndrome

In part two of this chapter I explain how an approach to conscious perception which respects *activity* and *knowledge-how* can be used to explain split-brain syndrome, and I do so by providing an RSE interpretation of the syndrome. I will argue that the RSE agent uses external factors (either bodily, environmental, or both) to act as a functional surrogate for the corpus callosum, and that this results in the brain's hemispheres making use of external factors to communicate with one another. I will then argue that this account can explain why perceptual consciousness in split-brain patients is unified in everyday life, and yet why a (temporary) split in consciousness occurs in experimental contexts. I explain that it can do so because it allows for external factors to play a constitutive role in perceptual experience. Finally, I provide a number of empirical predictions of the account and suggest some ways this proposal could be tested in an empirical setting.

### **§1: Interpretation Five- Experimental Aberration Model**

Scientists were surprised at the results of split-brain experiments because, aside from the experiments themselves, split-brain patients behave surprisingly normally. Consider the following comment by Springer and Deutsch:

The frequency with which these stories are mentioned makes it easy to forget that they describe rare incidents that are viewed as strange, isolated incidents even by the people involved. Thus, a pattern of sophisticated tests specifically designed to identify a commissurotomy patient would be needed for anyone to know the



operation had occurred. [Springer and Deutsch, quoted in Ferguson *et al.*, 1985, p. 503].

Split-brain patients rarely exhibit abnormal behaviour outside of hyper-constrained experimental conditions. Outside of the laboratory they are generally indistinguishable from healthy subjects. It is therefore not necessary to take the results of split-brain experiments to be representative of the general human condition, or even of split-brain patients themselves, outside of the laboratory. This consideration points toward a natural explanation of split-brain syndrome— split-brain patients possess a unified conscious perceptual field in everyday life which is disrupted only for the duration of the experiment.

Thomas Nagel rejects this *experimental aberration* account, and his reasons for doing so are worth quoting in full:

But the suggestion that a second mind is brought into existence only during experimental situations loses plausibility on reflection. First, it is entirely ad hoc, it proposes to explain one change in terms of another without suggesting any explanation of the second. There is nothing about the experimental situation that might be expected to produce a fundamental *internal change* in the patient. In fact it produces no *anatomical changes* and merely elicits a noteworthy set of symptoms. So unusual an event as a mind's popping in and out of existence would have to be explained by something more than its explanatory convenience.

But secondly, the behavioural evidence would not even be explained by this hypothesis, simply because the patient's integrated responses and their dissociated responses are not clearly separated in time. During the time of the experiments the

patient is functioning largely as if he were a single individual...The two halves of his brain cooperate completely except in regard to those very special inputs that reach them separately and differently. For these reasons hypothesis (5) [the experimental aberration model] does not seem to be a real option; if two minds are operating in the experimental 'situation', they must be operating largely in harmony although partly at odds. And if there are two minds then, why can there not be two minds operating essentially in parallel the rest of the time? [Nagel, 1971, p. 408, *my emphasis*].

If one looks closely at this quote (I have italicised the relevant parts) one can see that Nagel's argument is driven by an assumption of cognitivism. Nagel assumes that only internal brain-based differences are relevant to the structure of consciousness, and so he thinks the experimental aberration account is a non-starter.

In part two of this chapter I respond to Nagel's points. I argue that the *ad hoc* charge only applies to cognitivist theories, and I will argue that RSE predicts that there will be a difference between the conscious states of split-brain patients between every day and experimental scenarios. I conclude that RSE can explain split-brain syndrome in a parsimonious manner whereas cognitivism cannot. Therefore, RSE should be preferred as an account of perceptual consciousness.<sup>55</sup>

---

<sup>55</sup> Tim Bayne has recently objected to the experimental aberration model, and for similar reasons to Nagel: "But in fact the vehicle externalist proposal just sketched is really more of a promissory note than anything else, and it may not be one that is easily cashed...As far as I can see, the only plausible move to be made here appeals to the thought that the cognitive demands that split-brain patients face in experimental contexts are more taxing than those that everyday life places on them, and that it is these demands that govern the structure of the relevant causal flows...it is highly implausible to suppose that high cognitive load could generally lead to phenomenal division." [Bayne, 2010, pp. 203-204]. In the rest of this chapter I provide a fleshed out externalist account and, moreover, do so without arguing that "high cognitive load leads to phenomenal division" in the experiments.

## **§2: The Nomological Possibility of an Externalist Account of Split-brain Syndrome**

In the course of arguing against the coherence of the partial unity interpretation, Susan Hurley provided an externalist account of split-brain syndrome. Hurley did not intend to argue for the truth of her account. Rather, she argued for a weaker claim— that an externalist account of split-brain syndrome is a nomological possibility. In this section I outline her argument, and explain why it allows for the theoretical possibility of an experimental aberration account of split-brain syndrome which avoids Nagel’s *ad hoc* objection.

Hurley begins her argument by referring to people who have been born without a corpus callosum, who are known as “*acallosal* patients”. Acallosals appear to possess a unified conscious field, and they perform as well as healthy subjects during split-brain experiments. Consequently, the existence of acallosal patients demonstrates that it is nomologically possible for an organism to possess a unified field conscious field without possessing a corpus callosum.<sup>56</sup> Hurley provides an outline of two possible explanations as to how acallosals come to possess a unified conscious field [1998, p. 191]:

1. The left and right hemispheres of acallosals are connected to one another via alternative neural pathways.
2. The left and right hemispheres of acallosals use external factors (bodily, environmental, or both) as a functional surrogate for the missing corpus callosum.

---

<sup>56</sup>Acallosals do suffer from a number of psychological and developmental problems. For example, they possess a number of cognitive problems and often have difficulty engaging in social interaction. Their perceptual capacities, however, are no different from subjects with healthy brains.

Hurley admits that the most likely explanation as to why acallosals possess a unified conscious perceptual field is explanation (1). I agree with her on this point— (1) makes best sense of why acallosals perform the same as normal subjects in split-brain experiments and it is backed up with neurophysiological evidence (Tovar-Moll *et al.* [2014] have mapped the brains of acallosal patients and found connections in the mid- and fore- brains which are not found in subjects who possess an intact corpus callosum. They claim that these connections have developed as a response to, or to compensate for, the absent corpus callosum).

Although Hurley thinks that interpretation (1) provides the best explanation of acallosals, she does argue that interpretation (2) should be taken seriously as a nomological possibility. Her argument for this claim can be seen as a version of Andy Clark and David Chalmers' *parity principle*, which states that:

If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process.

[Clark & Chalmers, 1998, p. 8, *italics in original*]

Hurley argues that, if an acallosal did unify their perceptual field via external means, we would have no reason to deny that these external factors form a constitutive part of the acallosal's perceptual field. We accept that the corpus callosum plays an important functional role in unifying an organism's conscious perceptual states. It does so because it provides the means through which the left and right hemispheres can pass information to one another. Hurley argues that if the left and right hemispheres use the environment to pass information to one another in the same manner, then we have no reason to deny that these external factors are constitutive of the organism's perceptual field. To think

otherwise, according to Hurley, requires dogmatically assuming that there is a *magical membrane boundary* around the brain which insulates its consciousness providing properties from the inert, non-conscious properties of matter outside the brain.

Hurley argues that we have no reason to accept the idea that there is a magical membrane boundary surrounding the brain:

The causal relations between nervous systems and environments are intricate and continuous. There is nothing specially oomphy about causal relations inside the skin, or inside the head, nothing specially capable of pushing or shoving. So there is nothing causally mysterious or inhospitable to materialism or naturalism or realism about relational states of persons. And there is no magical causal boundary around persons. Viewed subpersonally, they are in principle transparent to causality.

[Hurley 1998, p. 336]

Hurley's statement in this passage is that, given the brain, the body, and the environment are all made of the same thing, we have no reason to a priori assume that there is something special about the brain which requires we treat it as if it possess a magical membrane, that insulates the consciousness providing properties of the brain from the non-conscious and inert matter in the body and the environment. Consequently, we have no naturalistically acceptable reason to reject interpretation (2) as a possible explanation for the unity of the acallosal's perceptual field.

Hurley develops this point by referencing a comment made by the psychologist Marcel Kinsbourne:

Kinsbourne remarks that the absence of the corpus callosum is biologically trivial, since minor adjustments in orientation distribute the same information to both sides, even in the absence of cross-cuing. [Hurley, 1998, p. 189]

By rejecting the magical membrane boundary Hurley concludes that external factors *could* play a constitutive role in unifying conscious perception. She takes Kinsbourne's comment to provide an example of *how* external factors could unify conscious perception— a tilt of the head can cause information to be passed into both hemispheres, and so this head-tilt can play the same functional role as the corpus callosum (which also ensures information is passed into both hemispheres). If we reject the magical membrane boundary, then we have no reason to deny that a split-brain subject which uses this method is unifying its perceptual states and that its bodily movements play a constitutive role in this perceptual unification.<sup>57</sup>

Hurley further strengthens her argument (for the nomological possibility of a conscious perceptual field unified via external factors) by introducing the *hypothetical acallosal* thought experiment. The hypothetical acallosal uses external factors to unify their perceptual states and they have done so since birth. Because external factors have played a constitutive role in unifying the acallosal's perceptual field since birth, Hurley assumes that only *ad hoc* or biologically chauvinistic considerations could motivate denying the hypothetical acallosal a unified conscious field. As such, she concludes that the hypothetical acallosal is nomologically possible.

---

<sup>57</sup>It may be objected that Kinsbourne's example merely shows that bodily movement can play the causal role of exposing each hemisphere to the same stimuli (as opposed to playing the constitutive role of unifying perceptual consciousness). Certainly, this does seem to be the point that Kinsbourne is making. For present purposes, however, it suffices that Hurley's construal is a nomological possibility.

Having argued it is nomologically possible that a human could use external factors to unify their conscious perceptual states, Hurley applies her argumentative strategy to the example of split-brain syndrome. She argues that immediately post-surgery the split-brain subject should not be seen as comparable to the hypothetical acallosal. If the split-brain subject does use external factors to pass information between its hemispheres at this stage, then they should be considered to possess two separate centres of consciousness (or one partially unified centre) which are working together.<sup>58</sup> According to Hurley, the two hemispheres of this split-brain subject are engaged in an explicit and personal-level co-ordination of their activities and are therefore metaphysically analogous to two agents working together. However, if the split-brain subject is able to subsume its use of external factors within automatic and sub-personal mechanisms (in the same manner as the hypothetical acallosal), then Hurley argues that we would have no principled reason to deny that external factors are playing a constitutive role in unifying the split-brain subject's perceptual field.

### **§3: From Possibility to Actuality— the Empirical Evidence**

I have just outlined Susan Hurley's argument that split-brain subjects *could* use external factors to unify their conscious perceptual field. I am now going to provide an empirical argument for the claim that split-brain subjects *do* use external factors to unify their conscious perceptual field. I begin by explaining the split-brain behaviour known as *cross-*

---

<sup>58</sup> Consider an analogy with two separate people who work in concert to perform a task. We would not consider these two people to possess a unified consciousness purely because they are functioning in concert with one another.

*cueing* (upon which much of my argument rests), before launching into the empirical argument proper.

### **§3.1: What is ‘cross-cueing’?**

*Cross-cueing* occurs when one hemisphere attempts to pass information to the other hemisphere by using external factors— the hemisphere uses external factors to ‘cue’ the opposite hemisphere. Consider as an example three common split-brain behaviours which have been observed by Joseph Bogen [1990]. In one of these behaviours, patients will manipulate a rubber band which has been placed in their left hand in order to communicate tactile information to the left hemisphere. The left hemisphere then becomes aware that the left hand is holding a rubber band. A second common behaviour involves subjects stabbing themselves with a pencil which has been placed in their left hand. This stabbing behaviour sends tactile information to the left hemisphere, which then becomes aware that the left hand is holding a pencil. A third commonly observed form of cross-cueing is known as *verbal cross-cueing*. In this form of cross-cueing, the speaking left hemisphere will voice verbal instructions to the right hemisphere, which allows the latter to complete the task which has been assigned to it by the experimenter.<sup>59</sup> The commonality of these behaviours is particularly striking when it is noted that “patients...only rarely if ever [meet] one another” [Bogen, 1990, p. 220]. Consequently, it seems as if split-brain patients independently engage in this kind of cross-cueing behaviour.

---

<sup>59</sup> Because verbal cross-cueing is so easily observed, it is generally noted by experimenters and measures will be put in place to eliminate it. For example, subjects are often told not to speak during experiments.



Interestingly, cross-cueing behaviour is not restricted solely to human subjects. In a 1969 paper, Michael Gazzaniga studied the evidence for the existence of cross-cueing behaviour in monkeys. The monkeys in question were able to engage in behaviour during split-brain experiments which was incompatible with the fact they had undergone split-brain surgery. Gazzaniga suspected that these monkeys had circumvented his experimental controls by tilting their heads to pass information between their anatomically disconnected hemispheres. He tested this hypothesis by fitting a restraining device to the monkeys which restricted their ability to move their heads. Once a restraining device was fitted to the monkeys' heads, their ability to circumvent the experimental design foundered. Gazzaniga concluded that the monkeys had been using movement to pass information between their hemispheres. Moreover, he claimed that this finding was applicable to a number of other studies which had been carried out on split-brain subjects:

The cross-cuing mechanism proposed appears sufficient to explain all previous reports of ipsilateral eye-hand control in cats, monkey, and man. [Gazzaniga, 1969, p. 16]

Many split-brain subjects engage spontaneously in cross-cueing behaviour. Even non-human animals, which presumably have little to no awareness of their peculiar situation, engage in this kind of behaviour. It is therefore plausible to think that cross-cueing is a universal behaviour which *can* occur without any explicit or deliberate intentional actions on behalf of a personal level agent. This provides some *prima facie* support for the idea that cross-cueing behaviour *can* be subsumed under sub-personal mechanisms.

### ***§3.2: Hurley's Criteria***

Having explained what cross-cueing behaviour is, and having noted that it could be applicable to the sub-personal level of explanation, we can return to the argument that split-brain patients do use external factors to unify their conscious perceptual states. Hurley claims that we can determine whether or not cross-cueing behaviour has been subsumed under one extended sub-personal system by using the following criteria [Hurley, 1998, pp. 192-193]:

1. If the use of external factors is accessible at the personal level and engaged in deliberately (for example, if a patient deliberately tilts their head in order to feed information to their opposite hemisphere), then this is evidence that two separate centres of consciousness are engaged in co-operation.
2. If the use of external factors is automatic and occurs without any conscious or personal level intervention, then their use can be taken to apply at the sub-personal level of explanation. For example, if a patient reflexively tilts their head without realising they have done so, then their head tilting can plausibly be construed as a sub-personal process. In such a case the split-brain patient's consciousness can be taken to have been unified by a sub-personal system which constitutively includes external factors (specifically, one which uses parts of the body or environment as functional surrogates for the missing corpus callosum).

It has already been noted that immediately post-surgery, if the two hemispheres do pass information to one another, they are best seen as metaphysically analogous to two co-operating personal level agents. However, if we can find evidence that such co-operation

between hemispheres occurs at the sub-personal level, then we can take external factors to be playing a constitutive role in unifying the split-brain patient's field of perceptual consciousness. These external factors would be fulfilling the exact same sub-personal functional role once fulfilled by the corpus callosum. Consequently, they should be accorded the status of being constitutive of perceptual experience even though they are located outside of the brain.

Although Hurley does offer these criteria, she does not explain how such a transition could occur. I am going to suggest that it occurs via the learning of new sensorimotor skills. I will then argue that we can empirically test whether split-brain patients develop these new sensorimotor skills by applying the *Dreyfus account of skill acquisition* to this scenario.

#### **§4: Split-Brain Syndrome and Sensorimotor Skill**

The reader will recall that RSE explains an organism's ability to perceive to be predicated on its possession and exercise of sensorimotor knowledge. It explains perception to be predicated on the possession and exercise of the *knowledge-how* required to become perceptually related to the environment. I want to suggest that we can explain how this *perceptual know-how* is acquired (in the case of split-brain syndrome) by applying the *Dreyfus account of skill acquisition* [Dreyfus & Dreyfus, 1980; Dreyfus, 2002]. If we apply this account to RSE, then we will be able to provide an operationalisation of the passage of sensorimotor knowledge from being applicable at the personal level of description to being applicable at the sub-personal level.<sup>60</sup> Consequently, we could determine whether split-

---

<sup>60</sup> Note, my argument does not rely specifically on Dreyfus' account of skill acquisition. It will work for any account of skill acquisition upon which skill: is taken to be a type of know-how; and, there is a continuum

brain patients do use sensorimotor skills which are partially constituted by external factors when they perceive.

The Dreyfus account distinguishes between five stages of skill acquisition which are labelled as follows— “*novice*”, “*advanced beginner*”, “*competence*”, “*expert*”, and “*mastery*”. Dreyfus’ account explains skill learning as initially applicable at the personal level of explanation. When we begin to learn a new skill we engage in deliberate personal level activity. However, as the learner passes through the various stages of skill-learning, their possession and exercise of the skill comes to apply more and more at the sub-personal level of explanation. Once a skill has been mastered, the Dreyfus account explains it to be explained solely in terms of sub-personal mechanisms. Thus, stage one is thought to apply at the personal level of description whilst stage five is thought to apply wholly at the sub-personal level. This schema can therefore be used to provide a means through which Hurley’s criteria can be operationalised and tested. I am going to argue that if we apply this account of skill acquisition to split-brain syndrome, we will find that split-brain patients exhibit evidence of learning a new sensorimotor skill which constitutively involves using external factors to replace the missing corpus callosum.<sup>61</sup>

---

between a beginner and expert, such that as a learner becomes more proficient their exercise and/or knowledge of the skill becomes more and more sub-personal. I use Dreyfus’ account in what follows because it is well-worked out, well-known, and has empirical support. However, I am not necessarily bound to the specifics of this account.

<sup>61</sup> Dreyfus’ account posits a continuum of skill-levels and so there may be occasions wherein determining the exact stage a learner is at will be entirely arbitrary. This will especially be the case in the examples I draw on in the next section, because there has been little-to-no empirical work carried out to test my specific argument. As such, the reader should bear in mind throughout that my placing of certain behaviours at a given skill-level is generally based on approximate estimations.

#### ***§4.1: Applying Dreyfus' Account of Skill to Split-brain Syndrome***<sup>62</sup>

At the first stage of skill acquisition, which Dreyfus labels the “*novice*” stage, the learner is given specific information about scenarios out of context and instructed to follow specific rules in these specific scenarios. The novice chess player, for example, will be informed of the numerical value of each piece and instructed to never engage in a negative exchange.<sup>63</sup> Novice learners will be unable to incorporate these instructions into the wider situational context of the activity in question, and so their knowledge is constituted by a series of autonomous, context-free, and rule-guided behaviours. For example, the novice chess player will be unable to recognise situations wherein a negative exchange would be appropriate. Rather, they will employ the rule ‘never engage in a negative exchange’ across the board when they play chess.

This stage would appear to best correspond to the state of the split-brain patient immediately post-surgery. If the patient does use external factors to pass information between hemispheres at this point in time, it would appear that the behaviour is engaged in deliberately. Although the patient is attempting to use external factors to replace the now-missing corpus callosum, their use of these factors would appear to be constituted by intentional and rule-guided behaviour. An example of cross-cueing which would appear to support this idea is an idiosyncratic behaviour carried out by patient N.G.:

[P]art way through test [N.G.] changed [her] manner of responding. Instead of pointing quickly with her hand, she would pause before her choices and move her

---

<sup>62</sup> The cross-cueing behaviours subsequently examined have all been taken from [Bogen, 1990], unless I explicitly state otherwise.

<sup>63</sup> A negative exchange occurs when the pieces lost by a player have a higher numerical value than the pieces taken from their opponent in a given exchange.

hand only after a motion of her head which resulted in her pointing of her chin toward the choice subsequently pointed out by the hand. When this was recognised and she was asked not to move her head she resumed the undelayed pointing with her right hand. [Bogen, 1990, p. 220]

N.G.'s behaviour appears to be a deliberate and personal level phenomenon because she would "pause before her choices". Furthermore, when asked to cease her behaviour N.G. was able to do so immediately (presumably, if the behaviour were automatic and sub-personal N.G. would have more trouble ceasing it). N.G. appears to have explicitly decided upon the use of this strategy in the specific context of the experiment she is subject too. Therefore, N.G.'s exercise of the skill 'use-external-factors-in-place-of-corpus-callosum' is best categorised at the novice stage of skill development.

The second *advanced beginner* stage of skill development is reached when the learner can recognise contextually defined situations and apply rules which are specific to that particular situation. For example, the advanced beginner will be capable of determining when an opponent's King's defence is weak and will be capable of applying the rule 'attack a weakly defended King' in light of their determination. Whether or not a King's defence is weak is something which can only be determined in the context of a specific game, and so the advanced beginner's decision to attack the King is based on situation-specific knowledge. The advanced beginner has moved beyond the novice stage because they are now capable of recognising context-specific situations and responding to them in an appropriate manner.

Split-brain patient R.M. shows signs of being at the advanced beginner stage of learning the skill 'use-external-factors-in-place-of-corpus-callosum'. R.M. exhibited a

number of (largely ineffective) cross-cueing behaviours when undergoing tests four months after surgery:<sup>64</sup>

When a pencil was placed in his left hand he held it appropriately but could not name it. It was then put into his right hand and he said, "a pencil." When a watch was put in his left hand, he said it was a "pencil" even when, with his left hand, he was holding the watch up to his left ear. A paper clip was put in his hand and he could not tell what it was; but when he put it in his right hand he immediately identified it. A pipe in the left hand was put into his mouth in an appropriate way; but it was called a "pencil" even after the bit was between his teeth. When an ashtray was put into his left hand he struck the table with it; it made a distinctive sound and he immediately told me what it was. When a pair of glasses was put in his hand, he could not name what he was holding until he tried to put them on. A handkerchief was put in his left hand; his left hand immediately put it into his left hip pocket...but he could not say what it was. [Bogen, 1990, p. 219-220].

R.M. is attempting to engage in cross-cueing behaviour and he does appear to recognise that, in the context of experimental situations, it is a good idea to use the skill 'use-external-factors-in-place-of-corpus-callosum'. However, R.M. is not very proficient in this skill and so his attempts to exercise it are largely unsuccessful.

At stage three of Dreyfus' hierarchy the learner is labelled "*proficient*". Proficient learners possess the ability to discern a number of task-relevant situations and they are capable of focusing on one task to the exclusion of the others. They will be able to act in

---

<sup>64</sup>I quote these behaviours at length because I will be returning to them shortly, in the context of discussing a different split-brain subject.

accordance with the actions appropriate for that particular task whilst ignoring the actions which could be undertaken were they to be engaged in a different task. For example, the proficient chess player will possess the ability to discern both that their opponent has a weak Kingside defence and that they themselves have a weak pawn structure. They will be capable of focusing and acting upon one of these situations to the exclusion of the other. If they decide to attack the opponent's King and that attack further weakens their pawn structure, they will be able to ignore the guideline 'rectify a weak pawn structure' in order to succeed in the task of attacking the opponent's King. Proficient learners follow *maxims* as opposed to situation-independent rules. Instead of prescribing specific actions for the completion of a given task, maxims prescribe general guidelines for context-dependent situations. The proficient chess player's attack on the opponent's King will be guided by the 'attacking the King' maxim, and the moves they make will be determined by the context of the particular state of the game. Their moves will not be guided by rules which prescribe specific moves which should be made when attacking the King.

The proficient stage of split-brain patients learning the skill 'use-external-factors-in-place-of-corpus-callosum' seems to best describe a set of behaviours observed by Kingstone and Gazzaniga [1995]. The word "o'clock" was localised to the subject's left hemisphere whilst the word "ten" was localised to its right hemisphere. When the split-brain subject was asked to draw what they had seen, they (surprisingly) drew a picture of a clock which was showing the time was ten o'clock. Kingstone and Gazzaniga performed a series of further experiments on this subject. They concluded that the subject had been able to circumvent the experimental conditions (which localised the words "ten" and "o'clock" to separate hemispheres) by allowing each hemisphere to control the subject's left hand. The



left hemisphere was initially given control of the left hand and it drew a clock. The right hemisphere was then given control of the left hand, and it drew the hands of the clock to indicate ten o'clock. Interestingly, Kingstone and Gazzaniga commented that "the only integration to be found here occurred 'on the sheet of paper in the drawing itself'" [1995, p. 324]. If my thesis is correct, such external factors do indeed play a constitutive role in integrating the subject's perceptual field.

Unfortunately, Kingstone and Gazzaniga did not investigate the extent to which the subject's co-ordination of hemispheric control of the hands was a deliberate and personal level decision or an automatic and sub-personal behaviour. So I can only provide a speculative explanation of the behaviour. That said, the behaviour is consistent with the subject having become used to applying the skill 'use-external-factors-in-place-of-corpus-callosum' even during experimental conditions. The subject's behaviour was novel in the context of split-brain experiments, so there is reason to believe the subject is proficient at the skill. My main reason for taking the subject to be only proficient at the skill (as opposed to more advanced) is that the subject's integrative behaviour was quite unimaginative. For example, when shown the words "hot" and "dog" the subject used the same technique (of switching hemispheric control of the hand) to draw a dog panting in heat. It would be more natural to interpret these words to refer to a piece of meat in a bun, and so a drawing of a sausage in a bun would be more appropriate. This indicates that the subject is not completely practised in applying the skill 'use-external-factors-in-place-of-corpus-callosum' during split-brain experiments. They are capable of applying the maxim 'switch-hemispheric-control-of-hands' during split-brain experiments. However, they are not particularly skilled at applying the maxim across different situations within the experiments.

Subjects who have reached the fourth stage of adaption are labelled as “*expert*”.

Experts possess the ability to ‘see’ salient aspects of a given situation, and they intuitively ‘know’ the best course of action to undertake without averting to deliberation of rules. An expert chess player, for example, will be able to ‘see’ that the best move to make is to mount an attack on the opponent’s King. They will not need to plan this attack by considering specific moves they should make. Rather, they just ‘know’ the best way to execute their attack.

Split-brain patient L.B. appears to be at the expert stage of learning the skill ‘use-external-factors-in-place-of-corpus-callosum’. Consider the following behaviour, which he carried out during a split-brain experiment:

After a picture of a tree appeared in the left hemi-field, his hands formed a triangle, and he then said ‘teepee’. [Bogen, 1990, p. 220]

L.B. appears to be well-practiced in using his body to cue his split-hemispheres. He appears to ‘see’ that he is in the context of a split-brain experiment and he appears to ‘know’ that the appropriate behaviour to engage in here is use his hands to ‘cue’ the opposite hemisphere. This kind of behaviour has only been observed in L.B. and so it seems he is at a particularly advanced stage of applying the skill ‘use-external-factors-in-place-of-corpus-callosum’. I therefore think the label expert is most appropriate for L.B.

The final and most advanced stage of skill learning is labelled “*mastery*”. People who master a skill exercise that skill in exactly the same manner as experts. The difference between an “expert” and a “master” is that the latter exercises their skill *absorbedly*. Absorbed agents still ‘see’ situations and ‘know’ how to react to them. However, they do so

automatically and lack any awareness of what they are doing.<sup>65</sup> Masters do not need to think about what they are doing. Indeed, (according to Dreyfus) thinking will actually have a negative effect on the actor's performance of the skill.

Dreyfus cites the following fact about chess players in support of the taxonomic term "master":

Excellent chess players can play at the rate of 5 to 10 seconds a move and even faster without any serious degradation in performance. At this speed they must depend almost entirely on intuition and hardly at all on analysis and comparison of alternatives. It has been estimated by cognitivists such as Herbert Simon that an expert chess player remembers roughly 50,000 types of positions. [Dreyfus, 2002]

There is absolutely no way that a master chess player could consult a data-base of 50,000 moves within 5-10 seconds:

Brains are extremely slow in comparison with turn-of-the-century desktop computers. Information is propagated down the axons of the nerves in the human brain at speeds of only a few meters per second— below the 55 mph speed limit! So a signal passing down an axon of length ten centimeters, say (quite a common long-distance connection within the brain), will take around a tenth of a second on its

---

<sup>65</sup> Although this is an anecdote I do take it to be relevant to the point at hand. I received a 'half-blue' award at my undergraduate institution, a sports scholarship from the 'Gaelic Player's Association', and played at the highest possible level in the Irish sport of hurling. I would therefore count as being an "expert" or "master" of hurling. Although this happened only occasionally, I did sometimes experience periods of time in games wherein it seemed to me as if I was performing actions automatically. For example, the ball would have been caught and struck down the field, by me, before I had even realised what I was doing. Such experiences would imply that at that time I was 'absorbed' in the game. Dreyfus motivates his account by focusing on and emphasising the phenomenology of skilled behaviour (although he does also use empirical evidence to support his view). I am attracted to his account largely because it fits in with my own phenomenological experiences (as well as that of a number of elite athletes with whom I have discussed such issues, cf. Csikszentmihalyi [2008])).

own, even before time is allowed for electronic spread within the dendrites, and for synaptic transmission. The signal propagation rate within the microchip that forms the central processing unit of a standard desktop computer is about one *million* times greater than this. [Carruthers, 2006, p. 25, *italics in original*]

Even *if* we grant that brains are computers, the limitations of their biological hardware entails that they cannot consult extremely large data-bases at the speed of man-made computers. It therefore seems highly unlikely that a chess master could consult a data-base of 50,000 moves in 5-10 seconds. Dreyfus argues that the best way to explain the behaviour of master chess players is to understand their behaviour to be wholly automatic, sub-personal, and to occur without thought. Master chess players do not need to think about what they are doing— they simply allow their body to perform the activity.

Split-brain patient C.C. appears to have achieved mastery of the skill ‘use-external-factors-in-place-of-corpus-callosum’. C.C. was asked to perform the same tasks of subject R.M. (quoted in detail earlier). However, he was much more effective in his execution of the tasks. C.C. was often observed to engage in non-verbal cross-cueing behaviours “which sometimes led to the correct answer and sometimes to an answer which was related but not correct” [Bogen, 1990, p. 219].<sup>66</sup> As Bogen notes, “it is quite instructive how an individual with a verbal IQ under 70 (65 at this time) can use minimal sensory cues to identify objects which are familiar to him.” [*ibid*, p. 219]. C.C. is not particularly well-equipped (cognitively) to actively decide to engage in cross-cueing. He is not cognitively well-equipped to successfully perform such behaviours. Finally, he does not appear to be

---

<sup>66</sup> His inability to correctly name some objects could be partially explained by his having a low IQ whilst being tested. For example, C.C. was unable to name a pine-cone even when exposed to one outside of test conditions.

deliberately engaged in their execution. Given these points, there is some reason to believe that C.C.'s cross-cueing behaviours are automatically deployed and so best described as applying to the sub-personal level of description. This is an attractive explanation of C.C.'s behaviour because C.C. was being tested eight years after surgery. We can therefore explain the difference between C.C.'s performance, and the performance of R.M. (who was tested only four months after surgery), by arguing that C.C. has achieved mastery of the sensorimotor skill 'use-external-factors-in-place-of-corpus-callosum'. C.C. excels in the skill 'use-external-factors-in-place-of-corpus-callosum' because he has practiced it for eight years. He can thereby be labelled a master of this skill.

Indeed, one could go further here and argue that the general behaviour of split-brain subjects in everyday life is best described in terms of their having mastered the skill 'use-external-factors-in-place-of-corpus-callosum'. Split-brain subjects are virtually indistinguishable from healthy controls in everyday life. Once they have adapted to the absence of the corpus callosum they primarily engage in aberrant behaviour only during experiments (however, even during experiments they still occasionally exercise the skill 'use-external-factors-in-place-of-corpus-callosum', as we have seen). Split-brain patients appear to deploy the skill 'use-external-factors-in-place-of-corpus-callosum' automatically, and the fact they are not aware that they do so indicates that this skill is being deployed absorbedly. Therefore, these subjects appear to display mastery of this skill. If subjects have mastered this skill, then it is best described as applying at the sub-personal level. The external factors used by the split-brain subject should therefore be considered to be constitutive of experience— they have replaced the corpus callosum and now play the functional role of integrating the anatomically split hemispheres.

### **§5: RSE and the Experimental Aberration Interpretation**

The reader will recall that the experimental aberration model takes the bizarre behaviour of split-brain subjects to be explained by the context of the experiment itself. Split-brain patients are thought to possess a unified perceptual field in everyday life and this field becomes split only during the course of split-brain experiments. The most popular objection to this account of split-brain syndrome is that it is *ad hoc*— there is no real, principled reason to think that engaging in split-brain experiments would cause a split in conscious experience.

The *ad hoc* objection is rooted in a cognitivist view of consciousness. As has been previously explained, cognitivists assume that consciousness is brain-bound because the representational vehicles of conscious states are situated within the brain. Engaging in split-brain experiments does nothing to change the anatomical or functional structure of the brain, and so it is argued that the experimental aberration model of split-brain syndrome is *ad hoc*— split-brain experiments do not change the brain's anatomical or functional structure, and the brain's anatomical or functional structure is thought to be constitutive of conscious experience. Therefore, there is no reason to think split-brain experiments would cause a change of structure in conscious experience. As such, cognitivists conclude that split-brain experiments reveal a fact about the conscious experience of split-brain subjects which is true of them even outside experimental conditions.

An account of consciousness which respects *activity* and *knowledge-how* can, however, provide a principled response to the *ad hoc* objection. In the case of RSE, the response would go as follows— split-brain experiments negate the subject's ability to exercise the skill 'use-external-factors-in-place-of-corpus-callosum' because they block off

the subject's ability to use external factors to unify their perceptual states. Given that these external factors act as a functional replacement for the corpus callosum, split-brain experiments are effectively equivalent to the subject under-going a second (temporary) split-brain procedure. Consequently, it should be expected that split-brain experiments would have the results that they do.

This is a particularly strong response to the *ad hoc* objection because it falls out naturally from the RSE account. RSE already takes external factors to be constitutive of perception, and it already understands perception to occur when an organism possesses and exercises sensorimotor knowledge, prior to consideration of split-brain syndrome. Each of these posits plays a crucial role in the RSE argument for the claim that split-brain patients learn the skill 'use-external-factors-in-place-of-corpus-callosum' after under-going split-brain surgery. Given that both externalism about perceptual consciousness and sensorimotor knowledge are invoked by RSE independently of split-brain syndrome, it therefore cannot be objected that RSE invokes these posits in order to accommodate this phenomenon. Indeed, rather than having to gerrymander its account of conscious perception to deal with split-brain syndrome, RSE arguably predicts the phenomenon—unlike acallosal patients, split-brain patients' brains lack the ability to develop new neural connections because they are too old. Consequently, it is likely that their brains will use external factors in place of the severed corpus callosum [*cf.* Anderson, 2014; Clark, 1997; 2008].

A further benefit of the RSE account of split-brain syndrome is that it can accommodate the partial unity and two-streams models of split-brain syndrome without requiring either account to be representative of split-brain experience at all times. Either of

these accounts (I am agnostic on which is to be preferred) can be applied to explain the conscious states of split-brain patients during experimental conditions, and each can also be applied to the split-brain patient during the early stages of their learning of the skill ‘use-external-factors-in-place-of-corpus-callosum’. However, once the patient has developed mastery of the skill, we can once more accord them the status of possessing a unified field of consciousness. By doing so we arrive at a more parsimonious account of split-brain consciousness, which avoids the conceptual and empirical problems facing the partial unity and two-streams models, without thereby rejecting the importantly correct aspects of these accounts (for example, that split-brain consciousness is not always unified).

### ***§5.1: Empirical Claims***

I have argued that accepting a view of consciousness which respects *activity* and *knowledge-how* provides one with the theoretical tools necessary to provide a parsimonious explanation of split-brain syndrome. I applied my RSE account to this phenomenon and argued that split-brain patients behave normally in everyday life because they exercise the skill ‘use-external-factors-in-place-of-corpus-callosum’. I suggested that applying Hubert Dreyfus’ theory of skill learning to this account could provide a means of operationalising the learning of this new sensorimotor skill. This is a novel argument, and no-one has investigated split-brain syndrome with this hypothesis in mind. Therefore, the empirical examples I used to support my account were necessarily speculative. The argument itself is, however, an empirical one. I can think of three ways that the account could be tested:

1. I have made the empirical claim that, over the course of learning the skill ‘use-external-factors-in-place-of-corpus-callosum’, personal level co-operative behaviour



eventually becomes subsumed under the sub-personal level. In the next chapter I will argue that sub-personal sensorimotor knowledge can be operationalised via (a non-representational version of) predictive processing. My theory could therefore be verified or falsified using the methods of the predictive processing paradigm. It predicts that there will be a marked difference between predictive neuronal activity immediately post-surgery and predictive neuronal activity post-adaption to the new sensorimotor skill— the predictive brain should exhibit lower activity levels when the skill is mastered, because there will be fewer errors being processed once it is mastered.

Although this claim would be difficult to test empirically, there is some evidence that brain activity does decrease when skill learning advances. For example, Gobel *et al.* [2011] have noted that brain activity decreases as subjects adapt to the learning of a novel skill. One could therefore apply a predictive processing version of Gobel *et al.*'s experimental protocol to my own hypothesis (about the split-brain subject's learning of the skill 'use-external-factors-in-place-of-corpus-callosum').

2. My argument could also be tested by focusing on the visual behaviour of split-brain subjects. If my argument is correct, split-brain subjects should exhibit noticeably different patterns of eye (and head) movement as they adapt to their newfound sensorimotor skill. This difference is predicted because, if external factors are used to pass information between hemispheres, then the subject must be using its perceptual capacities to pass the information between hemispheres. There should therefore be a noticeable increase in the visual behaviour of split-brain subjects after they have undergone the split-brain procedure.

It would in practice be difficult to perform this kind of test. However, one could attempt to do so by equipping subjects with a mobile eye-tracker and recording their eye (and head) movements as they navigate the environment. This experiment could feasibly be performed on macaque monkeys.<sup>67</sup> One could study the monkeys' perceptual behaviour pre-surgery and post-surgery,<sup>68</sup> perhaps by using Gazzaniga's 1969 experimental set-up as a starting point for our own. A marked difference between eye and head movement post-surgery would be required for my hypothesis to be empirically confirmed.

---

<sup>67</sup> I do not think I could condone the actual performance of this experiment on monkeys. However, my reasons for being unable to condone this experiment are ethical (not scientific) in nature.

<sup>68</sup> Importantly, because we have carried out the surgery, we will know exactly which parts of the monkey's brains have been severed. This is particularly relevant in the present experiment because most split-brain surgeries do not involve lesioning the optic chiasm. Lesioning the optic chiasm of the monkeys would ensure that visual information is not shared across hemispheres via neuroanatomical means.

3. Finally, my account relies on the empirical adequacy of both sensorimotor theory and Dreyfus' account of skill acquisition.<sup>69</sup> If either of these accounts are found to be empirically inadequate, then so too will my explanation of split-brain syndrome.

### §5.2: 'Siamese Twin' Objection

One could possibly respond to my argument by invoking the "*Siamese Twin Objection*". If my argument is correct it may turn out to have too strong a conclusion— two human agents who engage in extremely closely integrated behaviour could come to possess a unified field of perceptual consciousness. Michael Lockwood raises this objection by using the example of Siamese twins [1994]. Interestingly, this counter-example appears to be somewhat of a reality:

Figure Eleven: Shivanath and Sharam



[Ghosh, 2015]

<sup>69</sup> More specifically, any account of skill which takes skill acquisition to be a type of practical knowledge-how which becomes increasingly sub-personal as the agent becomes more skilled at the activity in question (*cf.* ft. 60).

Shivanath and Sharam can navigate their environment with relative ease and are capable of successfully co-ordinating their actions to complete everyday mundane behaviours (such as showering and dressing). Consequently, they could (presumably) co-ordinate their behaviour in such a way that their brains received sensory stimulation which accorded with Shivanath-and-Sharam-specific sensorimotor contingencies. If this is possible, then my account may appear to require that Shivanath and Sharam could come to possess a unified perceptual field in exactly the same manner as split-brain patients.

One could simply bite-the-bullet here and accept that Shivanath and Sharam could possess a single, unified conscious field. However, I reject this possibility and I do not think RSE allows for it. In order to explain why RSE does not entail this conclusion I must first distinguish between group *cognition* and group *consciousness*. Group cognition can be defined as groups of human beings bootstrapping their cognitive abilities to such an extent that the group of humans itself should be considered a minded entity. Group consciousness, however, requires that groups of agents could themselves possess a subjective point of view. There need not be ‘something it is like’ for a group of humans to constitute a single minded entity, but there must be ‘something it is like’ for a group of humans to constitute a single conscious entity. The Siamese twin objection is concerned solely with group consciousness,<sup>70</sup> and so to avoid this problem I need to rule out the possibility of group consciousness.

The reader will recall that my RSE account takes attention to be necessary for consciousness, and that attention itself is explained in terms of (a non-representational

---

<sup>70</sup> As such, I can be agnostic on the question of whether there are group minds.

version of) Cognitive Unison theory. Chris Mole has argued that groups of agents can exhibit Cognitive Unison, and so that there can be instances of group attention:

Sports teams are good examples of collectives that have and make use of understanding in this way, partly because they perform an activity that is *essentially* collective, and partly because the division of labour between different members of the team may make it the case that no individual member of the team understands the collective activity in its entirety. [Mole, 2011, p. 166]

His point is that, although individual members of a team may not understand how to perform a given task (such as win a match), the team as a whole can be considered to know-how to perform a given task (such as win a match). Consequently, it follows that the team can be considered to attend to the task at hand (of winning a match) whilst individual members of the team cannot. Therefore, the group itself can be ascribed the mental state of “attentively x-ing”. If Mole’s account of group attention applies to RSE, then it will follow that my own account falls prey to the Siamese Twin counter-example— if groups of agents can attend, then it would follow on RSE that group consciousness is possible.

I reject Mole’s claim that attention can be ascribed to groups of agents, and my reasons for rejecting this claim can be found in a paper by Adam Toon [2015]. Toon’s paper is focused on the topic of “group understanding”— he considers the proposition that scientific institutions should themselves be ascribed the predicate of “understanding (science)”, with the physical basis of this understanding spread between the various human agents working within the scientific institution and the equipment they use. Toon provides arguments from the extended mind literature to conclude that scientific institutions can be ascribed group-level understanding.

He also explains, however, that the conclusion that scientific institutions understand can be avoided if one explains understanding in terms of “an *ability to do* various things, such as apply theoretical principles to the world.” [*ibid.* p. 17, *my emphasis*]. Given that RSE is predicated on accepting the tenets of *activity* and *knowledge-how*, it does explain understanding (and therefore Cognitive Unison)<sup>71</sup> in terms of “an ability to do various things”. As such, the RSE construal of understanding allows for it to avoid the conclusion that groups of agents understand:

[One way to understand the ability view] construes understanding in (broadly speaking) behaviourist terms. On this view, understanding isn’t any underlying cognitive state or process; instead, to possess understanding is to exhibit the relevant sorts of behaviours, such as applying formulas successfully. On this reading of the ability view, it won’t make sense to ask whether understanding is internal or external (at least, not in the sense at stake in disputes over extended cognition).

[Toon, 2011, p. 15]

The *Extended Mind Hypothesis* is predicated on the assumption that cognition is representational, with the vehicles of mental representations themselves taken to be brain-bound. However, in certain conditions, it is argued that these representational vehicles can come to extend beyond the brain, because bodily or environmental factors can come to play the same functional role as internal factors. If one rejects a representational view of mentality, then it is no longer necessary to discuss the spatio-temporal location of representational vehicles (there are none). Mentality is instead taken to be constituted by certain capacities organisms possess which allow them to engage in certain activities.

---

<sup>71</sup>Cognitive Unison theory describes a specific type of understanding [Mole, 2011, ch. 4].

Accepting this capacity based view of mentality changes the dialectic of the extended mind debate— minds are considered to be *extensive* as opposed to extended [Hutto & Myin, 2013, ch. 8; Hutto, Myin, & Kirchoff, 2014]. On this construal of mentality, which is endorsed by RSE, “it won’t make sense to ask whether understanding is internal or external (at least, not in the sense at stake in disputes over extended cognition).”<sup>72</sup>

If RSE does not allow for group understanding, then it does not allow for group attention. If it does not allow for group attention, RSE does not allow for group consciousness. Therefore, RSE avoids the ‘Siamese Twin Objection’. RSE does not allow for the possibility that Shivanath and Sharam could possess a single, unified, conscious perceptual field.

## **§6: Conclusion**

In chapter four I argued that split-brain syndrome presents an insuperable problem for cognitivist accounts but that, by adopting a view of mind which respects *activity* and *knowledge-how*, one can provide a parsimonious account of the phenomenon. In part one I outlined the four possible cognitivist interpretations of split-brain syndrome and concluded that each was unsatisfactory. I explained that theorists have accepted these four interpretations largely because they are committed to a cognitivist framework. In part two I argued that accepting a view of mind which respects *activity* and *knowledge-how* allows one to accept the most parsimonious explanation of split-brain syndrome— that consciousness

---

<sup>72</sup> I am not here claiming that the various arguments for the extended mind thesis fail. Rather, I making a conditional argument of the following sort: if an ability view of mentality is correct, then the abilities in question cannot be considered extended. My reasons for preferring the antecedent of this argument are provided throughout the thesis.

comes to be split only during experimental circumstances. I explained that cognitivist accounts cannot accept this interpretation because it only works if factors external to the brain are allowed to play a constitutive role in unifying conscious perception. Then, I used my RSE framework to provide an *empirical*<sup>73</sup> argument for the claim that consciousness becomes unified in split-brain subjects when they learn the new sensorimotor skill ‘use-external-factors-in-place-of-corpus-callosum’.

### **§7: Postscript— Tye and Marks’ Argument**

I should note that there have been attempts by others to accept the experimental aberration interpretation. Charles Marks [1981] argues that split-brain subjects should be taken to possess a unified consciousness because their behaviour is best described (outside experimental contexts) in terms of a single functional system. In a similar vein, Michael Tye [2003, ch. 5] argues that the best explanation of split-brain behaviour in everyday life should lead us to conclude that consciousness only splits during the experiments. Although the specific arguments provided by each theorist for their positions differ, each is primarily motivated by the idea that the best explanation of split-brain research is that consciousness only splits during experimental circumstances.

---

<sup>73</sup> It is worth emphasising that the argument is empirical. Most arguments for the claim that perception constitutively includes external factors are motivated primarily on the basis of a priori, philosophical concerns. For example, such views are often motivated on the basis of phenomenological reflection [Martin, 2004; Strawson, 1979; Ward, 2012] or because they have positive epistemological consequences [Campbell, 2002; Mc Dowell, 1994]. My argument moves the debate away from its more esoteric philosophical concerns and into the realm of empirical science.

Tim Bayne objects to Tye's argument as follows:

Is this the best explanation of split-brain behavior? One issue on which Tye is noticeably silent concerns the mechanisms by means of which the patient's phenomenology is divided and then reunified. One might well wonder how the transition between everyday and experimental contexts could (dis)unify consciousness. Why might the split-brain patient's consciousness split into two 'under certain experimental conditions'? It seems plausible to suppose both that the structure of consciousness supervenes on neural structure, and that neural structure is not altered as the patient moves into and out of experimental conditions. [Bayne, 2005, p. 506].

Although Bayne does not directly address Marks' position, this objection could be applied to his account as well. Bayne argues that there is only one plausible way out of this bind for Tye (and Marks), and that is to accept a view upon which perceptual experience can constitutively involve external factors. Although not averse to externalist positions,<sup>74</sup> Tye has (to the best of my knowledge) not expressed sympathy with externalism about the vehicles of consciousness. Marks, on the other hand, advocates an explicitly Fodorian account of the mind. As such, it appears as if he cannot avail of an externalist theory of conscious perception [Fodor, 2009]. In any case, if Tye and Marks cannot accept an externalist view of mentality, their own accounts will fall afoul of the *ad hoc* objection. If they can, then they would be advancing accounts similar in spirit to my own (though the specific details of their accounts would differ substantially from mine).

---

<sup>74</sup>Tye is a well-known proponent of externalism about the content of representational mental states [1995; 2000]



## Interlude: The Benefits of *Activity* and *Knowledge-How*

In chapter three, I concluded that RSE should be preferred over its cognitivist rivals because it can explain conscious perception without facing the ‘hard problem of perceptual consciousness’ whereas cognitivist accounts cannot. In chapter four, I argued that RSE should be preferred over its cognitivist rivals because it can provide a parsimonious explanation of split-brain syndrome whereas cognitivist theories cannot. In this brief interlude, I am going to explain that RSE can arrive at these conclusions largely because it accepts the tenets of *activity* and *knowledge-how*.

The key aspect of RSE, which both underlies and drives the arguments presented in chapters three and four, is that it allows for the environment to play a constitutive role in conscious perception. RSE can allow for the environment to play such a constitutive role because it accepts the tenets of *activity* and *knowledge-how*. Activities themselves typically constitutively involve an interaction between an organism and its environment, whilst the exercise of knowledge-how is not the sort of thing that should be described as brain-bound. Consequently, acceptance of these two tenets naturally leads one to a view of mentality upon which:

‘[M]y mind’ does not stand for another organ. It signifies my ability and proneness to do certain things and not some piece of personal apparatus without which I could or would not do them. [Ryle, 1949/2000, p. 161]

A consequence of this capacity-based view of mentality which differentiates it from cognitivist theories is that one is not required to restrict mentality (in this case, conscious perception) to the brain. Rather, as Ryle explains:

The statement ‘the mind is its own place’, as theorists might construe it, is not true, for the mind is not even a metaphorical ‘place’. On the contrary, the chessboard, the platform, the scholar’s desk, the judge’s bench, the lorry-driver’s seat, the studio and the football field are among its places. [Ryle, 1949/2000, p. 50]

Theories which accept *activity* and *knowledge-how*, therefore, are not artificially constrained by the idea that conscious perception *must be* ‘contained’ or ‘located within’ the brain. It is for this reason that RSE is able to look beyond neural factors in order to both dissolve the ‘hard problem of perceptual consciousness’ and provide a parsimonious explanation of split-brain syndrome. Accordingly, we can see that acceptance of the tenets of *activity* and *knowledge-how* expands one’s theoretical options, and allows for theories which can account for philosophical and empirical puzzles which cognitivists simply have not been able to resolve. RSE arrives at the benefits detailed in chapters three and four largely because it respects *activity* and *knowledge-how*. This result gives us strong reasons to prefer theories like RSE over cognitivist rivals, and to pursue research in other areas of cognitive science which is guided by respect of these two tenets.

## Chapter Five: RSE and the Brain

A common criticism of approaches to mentality like RSE is that they cannot satisfactorily account for the brain's role in conscious perception. Although such approaches are generally considered to provide good phenomenological accounts of conscious perception which respect the fact that conscious organisms are embodied and embedded in a world, it is often argued that they provide only descriptions (as opposed to explanations) of conscious perception which ignore the undoubtedly key role played by the brain [Clark, 2009; Chemero, 2009, p. 93; Seth, 2014].

In this chapter I respond to this criticism and argue that it does not apply to RSE. I do so by focusing on the example of binocular rivalry. In section one, I explain what binocular rivalry is and explain that it is considered important by scientists largely because they have cognitivist commitments. In section two, I explain how the cognitivist theory of predictive processing proposes to explain binocular rivalry. Then, in section three, I explain how one could adopt this predictive processing explanation without invoking representation. Finally, in section four, I integrate this non-representational version of PP into RSE and highlight two benefits one accrues by accepting this integration. Consequently, I show that RSE can provide a mechanistic explanation of the brain's role in conscious perception. Furthermore, I argue that RSE improves upon cognitivist explanations of conscious perception because it can provide a phenomenologically plausible and naturalistically satisfactory explanation of empirical phenomena like rivalry, whereas cognitivist accounts cannot.

### **§1: What is Binocular Rivalry and Why is it Pursued as a Cognitivist Research Programme?**

*Binocular rivalry* occurs when each of a subject's eyes is concurrently presented with a different stimulus. For example, the left eye may be presented with a picture of a house whilst the right eye is presented with a picture of a face. It would be natural to assume that this experiment would result in the subject perceiving a mixture of the two stimuli— one might pre-theoretically expect the subject to visually experience aspects of both the house and the face meshed together. However, a meshing of the two stimuli is rarely experienced during rivalry. Instead, subjects usually experience one of the two stimuli at a given time, with the stimulus experienced constantly in flux. For example, the subject will perceive either the picture of the house or the picture of the face, with their experience constantly alternating between the two.<sup>75</sup>

Binocular rivalry has been intensively studied by psychologists because it is thought to be an ideal experiment for investigating the *neural correlates of consciousness* (NCC). The ultimate aim of the NCC research programme is to unmask “[t]he minimal set of neural conditions jointly sufficient for any one specific conscious percept” [Mormann & Koch, 2007]. Theorists working within this paradigm use the tools of empirical psychology (such as fMRI, single-cell recording, and so on) to discover the particular neural states which correlate with particular conscious experiences. Once enough correlates have been identified, researchers hope that they will then arrive at the neural *causes* of consciousness. Binocular rivalry has been studied by researchers within the NCC paradigm because it allows

---

<sup>75</sup> Subjects do *occasionally* experience a mixture of the two stimuli. This experience generally precedes, or occurs during, a switching of the stimuli. For example, a subject's experience of a house will at some point become fractured. Within the fractures of their experience certain aspects of a face will be experienced. The face stimulus will eventually be experienced as overwhelming the house stimulus— it comes to occupy more and more of the subject's visual field. Eventually, only the face stimulus will be perceived.

for the localisation of the neural correlates of conscious vision. In binocular rivalry experiments all factors are controlled for and held constant except for the experience being had by the subject. Scientists can therefore use rivalry experiments to study the changes in neural activity which correlate with the changes in visual experience, and in this manner can arrive at the neural correlates of visual experience (or so it is thought).

It should be self-evident that the NCC research programme is motivated by a cognitivist view of the mind— its explicit aim is to explain how the *brain alone* ‘gives rise’ to or ‘causes’ conscious experience. This view of consciousness can only be arrived at if one denies the tenets of *activity* and *knowledge-how*.<sup>76</sup> Although binocular rivalry has largely been pursued by cognitivist researchers for cognitivist reasons, I am going to explain how it can be accounted for within RSE. Therefore, I will show that RSE cannot be accused of ignoring the brain’s role in conscious experience. Moreover, I will then use the example of binocular rivalry to argue that RSE can in fact provide a better explanation of brain-based processing than cognitivist theories.

---

<sup>76</sup> To recap— *activity* is denied because conscious visual experience is thought to occur ‘inside the head’, whilst *knowledge-how* is denied because experience (or, the representational vehicles of experience) is (are) thought to be spatio-temporally located in (or constituted by) the brain [ch. 3, this thesis].

## **§2: PP and Binocular Rivalry**

Predictive processing has recently been applied to the example of binocular rivalry [Hohwy, Roepstorff, & Friston, 2008].<sup>77</sup> In this paper, Hohwy *et al.* argue that PP can both: explain *why* binocular rivalry occurs; and, explain apparently conflicting neurophysiological data.

### **§2.1: Why Does Rivalry Occur?**

Hohwy *et al.* argue that an explanation of why rivalry occurs requires a resolution to two problems— the *selection problem* and the *alteration problem*.

The *selection problem* is defined as follows:

[W]hy is there a perceptual decision to select one stimulus for perception rather than the other, and, further why is one of the two stimuli selected rather than some conjunction or blend of them? [Hohwy *et al.*, 2008, p. 690]

Hohwy *et al.* argue that the Bayesian theoretical apparatus inherent to PP provides the means through which to resolve this problem. The reader should recall that conscious perception on PP is thought to be constituted by the hypothesis which possesses the highest overall posterior probability.<sup>78</sup> In the case of binocular rivalry, there are three hypotheses

---

<sup>77</sup> Jakob Hohwy has more recently offered up a PP interpretation of binocular rivalry which “somewhat revises the proposal given in Hohwy *et al.* 2008” [2013, p. 22]. This updated interpretation is offered in the context of accounting for the unity of conscious experience, and is focused on allowing for action to play a more central role in the PP account of rivalry. I ignore this updated account in what follows. In this chapter my aim is simply to show how RSE could accept cognitivist accounts of certain phenomena, strip them of their representational baggage, and thereafter explain the brain’s role in conscious perception. Consequently, the exact nuances of a given PP account of rivalry are not of primary importance here.

<sup>78</sup> To recap— the posterior probability is the overall probability of a given hypothesis. It is calculated by multiplying the hypothesis’ prior probability and its likelihood. The prior probability is the probability given to the hypothesis independently of current events, whereas the likelihood is the probability given to the hypothesis in light of current events. For example, the prior probability of my being perceptually presented with a Snow Leopard is extremely low. Snow Leopards are very rare animals which live high in the Himalaya mountains, and few humans have ever cast their eyes on them. However, if I am on a Snow Leopard sight-seeing expedition in the Himalayas and am lucky enough to find a Snow Leopard in spite of its rarity, then the hypothesis with the highest likelihood will be that “there is a Snow Leopard in front of me”. The hypothesis

likely to be in contention as providing the best explanation of the visual sensory stimulation received by the brain:

1. *Face hypothesis* — the organism is being visually presented with a face stimulus.
2. *House hypothesis* — the organism is being visually presented with a house stimulus.
3. *Face-House hypothesis* — the organism is being visually presented with a face-house stimulus.

The prior probabilities for hypotheses (1) and (2) are going to be extremely high because human agents regularly encounter faces and houses in everyday life. Hypothesis (3), however, is going to be assigned an extremely low prior. Human beings rarely (if ever) encounter face-houses. Indeed, it is extremely unlikely that a human being would ever encounter two distinct objects inhabiting the exact same spatio-temporal location.

Consequently, the prior probability of (3) is going to be exceedingly low. Hypotheses (1) and (2) will be assigned a mid-to-low likelihood. Each of these hypotheses can account for (roughly) half of the visual sensory stimulation received, with the other half of the visual sensory stimulation received giving rise to error-signals for each hypothesis.<sup>79</sup> Hypothesis (3), however, will be assigned an extremely high likelihood. It can account for almost all of the visual sensory stimulation received by the brain.<sup>80</sup>

Given the foregoing, hypotheses (1) and (2) will both possess posteriors within a similar range. (3), conversely, will receive a posterior much, much lower than (1) or (2). The

---

“there is a Snow Leopard in front of me” will only dominate my PP hierarchy if the product of its prior and likelihood results in a posterior probability higher than other competing hypotheses.

<sup>79</sup> The reader should recall that error-signals occur when a given input differs from the input which was predicted. The brain’s sole aim, according to proponents of PP, is that of minimising prediction error. For this reason, the brain is constantly attempting to make its predictions as accurate as possible, and so the winning hypothesis will always be the one which has the least amount of prediction error.

<sup>80</sup> I say “almost all” because even the best predictions will result in some error.

reason for this is that hypothesis (3) has such a low prior that its posterior probability will never overcome that of (1) and (2). By averting to the Bayesian concept of “prior probability”, Hohwy *et al.* take themselves to have explained the selection problem. The reason that faces and houses are perceived in isolation during rivalry, whilst face-houses are not, is that the former have a prior probability which is exceedingly higher than the latter. Consequently, the strong priors of (1) and (2) mean that their posterior probability will always be higher than that of (3). Therefore, PP can explain why only faces or house are experienced during rivalry.

Having explained the selection problem via PP, Hohwy *et al.* then move on to tackle the *alteration problem*, which they define as follows:

[W]hy does perceptual inference alternate between the two stimuli rather than stick with the selected one? [Hohwy *et al.*, 2008, p. 690].

This is their explanation— the posteriors of both (1) and (2) will be quite close to one another (as we have seen). Consequently, there is little difference between the posteriors of both hypotheses, and so even a small amount of error from either hypothesis can result in the posterior of one eclipsing the posterior of the other. Hohwy *et al.* argue that the percept being experienced switches because the longer a given hypothesis is dominant the more likely it is to accumulate error. When the dominant hypothesis accumulates enough error, its posterior will become lower than the previously non-dominant hypothesis, and at this point the latter will be experienced. Hohwy *et al.* suggest two mechanisms which could be used to explain the accumulation of error— the influence of priors and the influence of *stochastic resonance*.



Hohwy *et al.* hypothesise that the brain (as PP construes it) may attach the prior probability that “the environment is constantly in flux” to almost all of its perceptual hypotheses. If the brain does attach this prior to its perceptual hypotheses then, in rivalry, the longer a given percept dominates the more error it will generate. As time goes on, the dominant hypothesis’ posterior will lower because it begins to violate the prior which has been attached to it (that “the environment is constantly in flux”). Alternatively, Hohwy *et al.* suggest that *stochastic resonance* results in alteration of the stimulus experienced because it results in “random change, due to neuronal noise, in the brain’s state” [Hohwy *et al.*, 2008, p. 693]. Neuronal noise could result in a change between hypotheses because it alters their posterior probability, such that the posterior of one hypothesis becomes lower than another, and so the hypothesis currently dominating the neural hierarchy will change. Hohwy *et al.* also note that alteration could be explained to occur via a mixture of the influence of priors and stochastic resonance. The important point here is that, once again, the theoretical apparatus of PP can be used to explain binocular rivalry (because it explains the alteration problem).

In sum, Hohwy *et al.* explain the selection problem with reference to the influence of priors on the overall posterior probability of a given hypothesis. They explain the alteration problem by referring to the influence of priors and stochastic resonance. Having explained both of these problems, they therefore take themselves to have provided a PP explanation as to why binocular rivalry occurs.

## **§2.2: How does PP account for the empirical evidence?**

In addition to explaining why rivalry occurs, Hohwy *et al.* think that PP can provide a strong explanation of the empirical data collected on binocular rivalry. One of the chief virtues of a PP account of rivalry, they argue, is that it can explain apparently conflicting empirical data within the one framework.<sup>81</sup>

It is well-known that neurons have their own *preferred stimulus*. For example, some neurons react strongly to the presentation of faces whereas others will not respond at all to faces. On bottom-up frameworks, these neurons are thought to record the presence of their preferred stimulus in the sensory array. PP frameworks, however, conceptualise these neurons as *expectation neurons*. For example, a neuron which reacts strongly to the presence of faces is to be understood as a face expectation neuron. It does not record the presence of a face. Rather, it records an expectation (or a prediction) that a face will be present in the sensory stimulus [Clark, 2013, p. 190].

In *functional magnetic resonance imaging* (fMRI) the brain is scanned and changes in blood-flow to its various areas are tracked. When a given area of the brain is active during a cognitive activity it will receive an increased amount of oxygenated blood. Consequently, fMRI is used to track changes in blood-flow because these changes are thought to reflect the neural activity of a given brain area. *Single-cell recording*, as the name suggests, involves the use of microelectrodes to record the activity of single neurons within the brain.

---

<sup>81</sup> Hohwy *et al.* provide an in depth explanation of how PP accounts for much of the empirical data collected on rivalry, but I will not rehearse these points here. Instead, I will focus on explicating how PP is supposed to account for apparently conflicting empirical data within the one framework. The ability to account for this conflicting data is supposed to be the major empirical virtue of PP accounts of rivalry.

The data collected from fMRI scans of the brain during rivalry and the data collected from single-cell recording is often thought to conflict. fMRI scans show activity at all levels of the visual system during dominance of those area's particular preferred stimulus. Single-cell recording, however, appears to show that only 'higher' levels of the visual system are active during the dominance of their preferred stimulus. The results of fMRI and single-cell recording thus appear to be contradictory— fMRI records activity in all levels of the visual hierarchy during dominance of the preferred stimulus of those areas, whereas single-cell recording does not.

Although these empirical results have proven problematic to account for, Hohwy *et al.* argue that PP has the theoretical resources to reconcile this (seemingly divergent) data within the one framework. They contend that:

fMRI correlates of rivalry may be driven by top-down predictions, whereas electrophysiological responses may reflect predictions or prediction error, depending on which population or unit is recorded. [Hohwy *et al.*, 2008, p. 698]

Their suggestion is that the fMRI data can be explained as tracking the activity of predictions as they make their way down the neural hierarchy. PP does posit that prediction occurs at all levels of the hierarchy, and so one would expect there to be evidence of prediction at all levels of the brain during dominance of a given preferred stimulus. Where single-cell recording is concerned, although there is covariation between activity in neural areas and their preferred stimulus in higher levels of the brain's hierarchy, single-cell recording in low- and mid- levels of the brain sometimes records evidence of activity during dominance, and sometimes records activity during periods of non-dominance, for a given neural area's preferred stimulus. Hohwy *et al.* argue that this neural evidence serves to support a PP

interpretation of rivalry. One would expect synchrony between cell activity in higher-levels of the brain's hierarchy and presence of their preferred stimulus because the winning prediction is issued from these levels. Conversely, one would not expect to record neural activity only during dominance of a given neuron's preferred stimulus at low- and mid-levels of the brain because the cell in question may be firing because it is signalling error or because it is passing on a prediction. Thus, single-cell recordings can appear to conflict with fMRI data because the former can be picking-up on either prediction- or error- signals, whereas the latter picks-up only on prediction-signals. By introducing the theoretical apparatus of PP, Hohwy *et al.* therefore take themselves to have resolved an apparent conflict in the empirical data.

In short, Hohwy *et al.* propose a PP account which explains why rivalry happens, explains the empirical evidence collected on rivalry, and which can even reconcile some apparently conflicting empirical work within the one framework. For this reason, they conclude that PP provides a strong conceptual framework for understanding binocular rivalry (and, therefore, conscious visual perception).

### **§3: Non-Representational PP**

In section four, I will explain how PP can be subsumed within RSE. In order to subsume PP within RSE, however, I must be able to provide an account of PP which does not violate the tenets of *activity* or *knowledge-how*. In this section I will provide such an account by arguing that the key PP posits from section two ("prior", "prediction", and so on) can be understood without requiring the concept of "representation". I will do so by drawing on arguments in this vein recently proffered by Nico Orlandi [2015; *cf.* 2014] and explicitly linking them in

with William Ramsey's *job-description challenge*. I will argue that these PP posits do not meet the job-description challenge, and so conclude that the PP explanation of rivalry can be accepted without requiring the adoption of representation.

### **§3.1: Ramsey's "Job-Description Challenge"**

The concept "representation" is a psychological notion, which requires the invocation of properties such as "aboutness", "truth-conditions", and "content". These properties are not typically thought to exist in non-cognitive physical entities, such as stones or trees. As such, they are to be considered special, higher-level properties, which are not ubiquitous in the natural world. Accordingly, we must have good reasons to describe a given mechanism or system in terms of representation.

In his 2009 *Representation Reconsidered*, William Ramsey proposes what he labels as "*the job-description challenge*". He argues that this challenge must be met if a given mechanism or system is to be described in terms of representation. In order to pass this challenge, Ramsey argues that the mechanism must: (1) play a functional role within a system which we would pre-theoretically regard as representational; and, (2) it must be explanatorily beneficial to treat the mechanism as functioning in this manner. Ramsey argues that condition (1) must be met because, otherwise, the concept of "representation" would become empirically vacuous. If we re-define the concept "representation", such that our use of the concept has nothing in common with our pre-theoretical psychological notion, then ascription of this concept would become meaningless. If condition (1) is not met, the "representational theory of mind" becomes a representational theory in name

only. Condition (2) must be met, according to Ramsey, because otherwise the concept “representation” would become explanatorily vacuous. We can trivially describe any mechanism or system in terms of representation. For example, I can describe a stone-rolling down a hill in terms of the stone *desiring that* it reaches the bottom, *believing that* rolling is the best way to achieve this aim, and so on. However, we do not receive any additional explanatory benefit from this representational account of stone rolling over and above that received from applying a purely non-representational, physicist’s account to the stone’s rolling behaviour. In contrast, describing my typing of this thesis in terms of representation is explanatorily beneficial. We can gain a better understanding of my typing this document if we describe my behaviour in terms of representation— I am typing this thesis because I *desire* a PhD in philosophy, and I *believe that* I must write a thesis in order to receive this qualification— as opposed to providing a purely non-representational, physicist’s description of my behaviour. Thus, condition (2) must be met in order to avoid a complete trivialisation of the concept “representation”. Representation should only be ascribed to a mechanism if ascribing representation helps one gain a better understanding of that mechanism.

In short, “representation” is a psychological concept which is not ubiquitous throughout physical systems in nature. If we are to maintain a robust and empirically useful notion of representation then, in Ramsey’s view, a given mechanism should only be described in terms of representation if it deserves to be described in representational terms. He proposes that we assess whether a mechanism does deserve to be described in terms of representation by submitting it to the job-description challenge. Ramsey argues

that a mechanism should be described in terms of representation only if it passes this challenge.

### **§3.3: “Prediction-” and “Error-” Signals**

Let us now submit the PP concepts of “prediction-signal” and “error-signal” to the job-description challenge. Both of these concepts appear *prima-facie* to require representation— predictions signal *that* such-and-such is the case, whilst errors signal *that* such-and-such is not the case. Nico Orlandi argues, however, that closer inspection of the role these terms play in PP explanations reveals they are not representational posits. They fail the job-description challenge.

Predictions in low- and mid- levels of the perceptual hierarchy serve to pass on predictions to the level *immediately* below them. Error-signals, on the other hand, pass error up to the level *immediately* above them. As such, each is concerned only with proximal conditions. Therefore, the signals being passed up and down the low- and mid- levels of the perceptual hierarchy are better understood in terms of causal covariation or correlation. In order to argue that these signals are representational in nature, one must therefore explain how or why brain-based causal covariation results in or requires representation. William Ramsey argues that accounts of brain-based representation founded upon causal covariation fail the job-description challenge, and so he concludes that mechanisms in the brain which function on the basis of causal covariation should not be considered representational [2009, ch. 4]. His summary of this conclusion is worth quoting in full:

Despite its common appeal, the receptor notion of representation [Ramsey's name for covariation based accounts] comes with a job description that, in this context, has little to do with the role of representation... When we look at the role of receptors *inside* of cognitive systems, as described by cognitive theories that employ them, we see that the role is better described as something like a reliable causal mediator or relay circuit which, as such, is not representational in nature. In other words, when a causal/physical system (like the brain) is described as performing various cognitive tasks by employing a structure that has the job of causing something to occur when and only when something else occurs, then the system is not, on the basis of this description alone, employing internal representations.

[Ramsey, 2009, p. 149, *italics in original*].

Ramsey's argument, in essence, is that we gain no extra explanatory purchase by treating causal covariation within the brain in terms of representation. We do not arrive at better accounts of neural processing by giving causal covariation between neurons a representational status, because treating them as such does not provide one with any extra explanatory benefits over those one would accrue by treating them as mere non-representational causal correlations.

Consider, as an example, the water fountain in my back garden which is powered by a small solar panel. The fountain only runs when sun shines on the solar panel, and so the running of water in this fountain causally co-varies with the presence of sunlight. Although we could treat the correlation between the running of water and the presence of sunlight in terms of representation (saying that running water represents sunlight), doing so does not provide one with any explanatory benefit over and above treating the running of water and



sunlight presence in terms of simple causal mediation (sunlight causes the running of water). Indeed, treating this series of events as involving representation would run directly counter to our ordinary use of the term “representation”. Ramsey’s argument is that causal covariation between neural processes is no different from causal covariation between sunlight presence and the running of water. Just as it is the case with the water fountain, we gain no explanatory benefit from treating causal covariation between neural processing in terms of representation, and doing so violates our pre-theoretic use of the concept. Therefore, causal correlation between neural processes fails the job-description challenge, and so does not deserve a representational status.

The proponent of representation is likely to object, however, that there is an important difference between the causal covariation between running water in fountains and sunlight presence, and the causal covariation present in neural processing. Consider, for example, Andy Clark and Josefa Toribio on this point:

The universe is stuffed with correlations and it is implausible to count them all as representations (think of accidental correlations). We agree, but note that the correlations between, for example, specific brain states and color perception look to fall onto the intuitively acceptable side of such a divide. [Clark & Toribio, 1994, p. 417]

A proponent of this kind of argument will agree that causal covariation, taken alone, is not sufficient for representation. However, if this causal covariation occurs in a biological organ (like the brain) and has proven evolutionary beneficial (has been selected for by the forces of natural selection), then it should be described in terms of representation.

One could, for example, argue for a teleosemantic account, upon which representation is thought to occur when there is causal covariation which has been selected for by evolution because it performs a fitness enhancing role [Dretske, 1988; Millikan, 1984]. On this kind of account, we determine the representational function of a given mechanism by averting to its evolutionary history. Neural states which co-vary with a given  $x$  would be taken to represent  $x$  because they were selected for by evolution to respond to it. If the neural states happen to co-vary with  $y$ , where  $y$  is a non-natural stimulus introduced in the lab, then the neural states will misrepresent because natural selection did not select them for signalling  $y$ . Teleosemantics requires ascribing representation by determining the function of a given mechanism, with this function determined in turn by considering what the mechanism itself was selected to do by the forces of natural selection. We could therefore argue that covariation between neural processing is representational because it has been selected for by the biological forces of natural selection, whereas covariation between sunlight presence and running water is not representational because it has not been selected for by evolutionary forces.

Buttressing the concepts of “prediction-signal” and “error-signal” with teleosemantics, however, is not going to help these concepts pass the job-description challenge. Both prediction- and error- signals are concerned only with proximal conditions, and the functioning of both can therefore be adequately accounted for without invoking the concept “representation”. Consequently, applying teleosemantics to this particular example will not help with the job-description challenge, because teleosemantics is concerned primarily with the *content* of a given representation. Applying teleosemantics to a given mechanism involves discerning when representation and mis-representation occur. As such,

teleosemantics is only applicable to scenarios in which the concept of “representation” has already been applied— the application of teleosemantics to a mechanism requires the assumption that a given instance of causal covariation is representational, and thereafter attempts to naturalise the particular content of that representation. The theory will not, therefore, help one in determining whether or not a given mechanism deserves to be described in terms of representation to begin with. In short, the theory of teleosemantics is only applicable once a given mechanism has already passed the job-description challenge. Consequently, teleosemantics cannot be used in an argument for the claim that instances of causal covariation pass the job-description challenge [*cf.* Hutto & Myin, 2013].

In sum, the concepts “prediction” and “error” should be understood to involve mere causal correlation between neural processes. Causal correlation should not be considered sufficient for representation because it fails the job-description challenge. We do not gain any extra explanatory purchase by treating the causal covariation entailed by “prediction-signals” and “error-signals” in terms of representation. Although one can buttress causal correlation with concepts from evolutionary biology, doing so will not help with the job-description challenge. Therefore, I conclude that we can satisfactorily account for the role of low- and mid- level prediction- and error- signals within the brain by conceiving of them in terms of mere causal mediation.

### ***§3.3: “Priors” and “Likelihoods” Are Non-Representational***

Having outlined Orlandi’s argument that prediction- and error- signals fail the job-description challenge, I am now going to explain why Orlandi thinks that the concepts

“prior” and “likelihood” (which are thought to be applicable at higher levels of the neural hierarchy) also fail the job-description challenge. Orlandi argues that these concepts are best understood as referring to non-representational biases present in the neuronal system:

Understanding perceptual priors, hyperpriors and likelihoods as biases means thinking that, as a result of repeated exposure to the structure of the world in the evolutionary past *and* in the present, the perceptual system is skewed to treat certain stimuli in a certain way. [Orlandi, 2015, p. 25, *italics in original*]

She argues that theorists are tempted to explain biases in terms of representation largely because they are in the grips of the traditional cognitivist idea that perception is to be understood in terms of internal inferential transitions between premises and conclusions in some kind of *language of thought* [cf. Ramsey, 2009]. She argues, however, that biases are more realistically understood to fulfil “the simple function of marking a hypothesis as more or less probable. They are like valves. They skew the brain toward certain neuronal arrangements” [Orlandi, 2015, p. 25].

Consider, once more, the small water fountain in my back garden. When the water reaches the bottom bowl of the fountain, it pools around the opening to a small pump. This pump pushes a small amount of this water back up to the top of the fountain, where the water once more begins to trickle down. Only a small amount of water is pumped up to the top of the fountain at any given time, because if too much were to flow from the top at once the fountain would quickly overflow. The pump therefore plays the functional role of biasing the flow of a system of water, such that there will always be a large pool of water at the bottom of the fountain and only a small amount at the top. We would not, of course, describe the biasing function of this pump in terms of representation. The main reason for

this, presumably, is that the ascription of representation to this pump-system would fail the job-description challenge. Describing the pump's biasing role in terms of representation does not provide one with any explanatory benefits over and above those one would accrue by simply treating it as a mere non-representational bias in a system.

Orlandi contends that a similar conclusion should be drawn in the case of the PP concepts "prior" and "likelihood". She argues that these concepts should be taken to refer to certain biases within a neuronal system, and that one should not treat these biases in terms of representation because there is no explanatory benefit in doing so. Consequently, she concludes that the concepts "prior" and "likelihood" should be taken to refer to mechanisms in the brain which play the non-representational function of pre-disposing neural systems to enter into certain arrangements whenever presented with a given environmental stimulus.<sup>82</sup> Once more, the argument I am presenting here is not that biases cannot be understood in terms of representation. Rather, it is that their functioning can be understood entirely without invoking representation, and that treating them in representational terms is not explanatorily beneficial. Biases do not pass the job-description challenge, and so we have no reason to treat them in terms of representation. Therefore, I

---

<sup>82</sup> Orlandi provides a number of other arguments against understanding biases in terms of representation. She concludes that treating biases in representational terms is misguided because biasing mechanisms lack a number of characteristics which she takes to be necessary for a given mechanism to count as representational (namely, *de-coupleability* and *availability for use in reasoning and the planning of action*). Furthermore, she motivates her non-representational stance toward such PP posits by noting that, by accepting a non-representational conception of priors, one need not face the problem of explaining where the priors come from. Bayesian frameworks all require an initial set of priors to get the Bayesian process going. However, the priors themselves and their values occur (more or less) *apropos of nothing*. Orlandi claims that her non-representational understanding of priors is preferable because it is much easier to explain how biases could arise via interactions between an organism and its environment, than it is to explain how representational knowledge could be accrued by brains. Although I am sympathetic to Orlandi's arguments, I do not focus on them here because defending them would require straying too far from present concerns.

conclude that the PP concepts “prior” and “likelihood” should be taken to describe non-representational biasing processes occurring within the neural system.<sup>83</sup>

### ***§3.3: “Posterior Probability” is Non-Representational***

Although Orlandi argues for a non-representational understanding of PP processing, she concludes that the results of this processing (the resulting ‘winning hypothesis’) should be understood in terms of representation. Orlandi arrives at this conclusion because she thinks that the winning hypothesis fulfils her three criteria for representation:

[R]epresentations are only those performance-guiding structures that are de-coupled from their causes, where this fact materialises in their standing for distal or absent conditions. [Orlandi, 2014, p. 133].

According to PP, the hypothesis with the highest posterior probability will determine the nature of our experience. Orlandi claims that this winning hypothesis is concerned with distal conditions because it is formulated on the basis of sensory information received by the brain (photons, sound-waves, and so on) and yet is itself about things beyond brain-based sensory receptors (such as laptop speakers). She argues that the winning hypothesis is de-coupleable from its environmental causes because it can be deployed even in the

---

<sup>83</sup> This non-representational construal of PP biases can be studied and assessed via empirical means. Orlandi suggests that we study the creation and maintenance of such biases via the area of visual science known as “*visual scene statistics*” [Orlandi, 2014; 2015]. In later sections of this chapter, I will suggest that we can study how high-level aspects of PP function as non-representational neuronal biases (in the case of binocular rivalry) by linking PP more closely with Anderson’s non-representational biased-affordance-competition framework. RSE itself is compatible with, and can accommodate work carried out in, both the visual scene statistics and biased-affordance-competition frameworks.

absence of its environmental causes.<sup>84</sup> Finally, she argues that the winning hypothesis deserves to be treated in terms of representation because it is used by the brain to reason with and plan action. Consequently, according to Orlandi, although PP processing itself does not deserve a representational status the result of this processing does.

In the context of RSE, however, the winning hypothesis does not fulfil these three conditions (or so I contend). The winning hypothesis in the RSE framework is identified with the affordance currently in charge of determining the organism's behaviour.<sup>85</sup> Affordances are personal level phenomena which are directly perceived by the organism itself. As such, they concern proximal (and not distal) states of affairs. Therefore, in the context of RSE, the winning hypothesis does not fulfil Orlandi's first condition of representation.

The direct perception of affordances also requires that, on RSE, the winning-hypothesis is not considered to be de-coupleable. Direct perception of affordances (and therefore environmental objects) is entirely constituted by what Anthony Chemero labels a "tracking relation". Tracking occurs when there "is...constant causal connection" [Chemero, 2009, p. 57] between the perceiver and perceived object. This relation is most obviously instantiated when an organism is directly confronted with an environmental object, in which case it is labelled "*effective tracking*". However, the relation can be maintained even when the object itself (temporarily) disappears from view. Such *non-effective tracking* of a (momentarily) hidden object "could be accomplished just by causal connection." [*ibid.* p. 57]. For example, if I am viewing a fox, I can continue to track it even if it disappears behind a rock by continuing to move my head and eyes at a similar momentum to the fox, such that

---

<sup>84</sup> Orlandi is happy to admit that visual perception may only be partially de-coupleable. I do not wish to dispute this aspect of her argument here. However, the interested reader can consult [Gładziejewski, 2016], for an interesting discussion of whether visual perception in PP is to be understood as partially or fully de-coupleable.

<sup>85</sup> I explain this point in more detail in section four of this chapter.

I follow the fox's trajectory even though it is itself momentarily out of view. In either case, there is a causal connection between the perceiving organism and the object being perceived. Therefore, direct perception is not de-coupleable. Consequently, on RSE the winning-hypothesis is not considered de-coupleable. It is identified with the winning-affordance, with the winning-affordance itself considered to be directly perceived.

This leaves us with Orlandi's final condition, upon which the winning hypothesis should be described in terms of representation because it is used for thought and the planning of action. RSE defines consciousness in terms of *access consciousness* (availability for use in thought and action). Access consciousness itself is thought to be instantiated via attention, which itself is empirically implemented through biased-affordance-competition. Biased-affordance-competition requires that the winning affordance will determine what the organism is conscious of, and so the winning affordance in the RSE framework has the same properties as the winning hypothesis in the PP framework. If Orlandi's reasoning is sound, then RSE is required to take consciousness to be representational.

This conclusion is implausible. I am inclined to reject it because it appears to be based upon what William Ramsey [2015] has labelled the "*representation demarcation thesis*" (RDT). Ramsey defines RDT as "the view that cognitive processes necessarily involve inner representations and cognitive theories must thereby be about the representational states and processes" [2015, p. 4]. Orlandi's argument for ascribing a representational status to the winning hypothesis appears to be based on acceptance of a version of the RDT because, by her reasoning, it is a *conceptual truth* that any states which involve the planning and execution of action are representational states.



Ramsey provides three arguments against RDT. The first reason he provides for rejecting RDT is that it requires a conceptualisation of cognition which is arrived at by largely a priori means. Obviously, whether or not a given instance of cognition is to be understood in terms of representation should primarily be an empirical matter. However, by defining cognition in terms of representation (as RDT does), one guarantees that no instance of cognition will ever be non-representational. Either we must find a representational explanation of the cognitive activity in question, or, it will not count as a cognitive activity at all. Ramsey thus rejects RDT because it requires cognitive science accept a priori constraints on its domain of study and he thinks that no serious science should accept such constraints.

Ramsey's second reason for rejecting RDT is that RDT undermines the empirical status of the *representational theory of mind*. As we saw in chapter one, the concept of "representation" is deployed by cognitivists because it is supposed to perform an indispensable role in explaining how organisms cognise in spite of facing a poverty of cognitive stimulus. It is therefore proposed as a theoretical posit which is supposed to play an empirical role in providing an empirical explanation of cognition. If, however, one accepts RDT, the cognitivist research programme within cognitive science no longer looks to be empirical in nature. Representation is not being posited for empirical reasons, because it plays an important role in an empirical theory of cognition. Rather, it is proposed for conceptual reasons, because it is a priori assumed that any empirical theory of cognition must be a theory of representation. Representation thereby becomes an unfalsifiable theoretical posit, and so the cognitivist research programme loses its empirical credentials. Consequently, Ramsey's second reason for rejecting RDT is that it requires an unscientific approach to cognitive science.

Finally, Ramsey rejects RDT because it encourages a wildly deflationary understanding of representation, such that even mere causal mediation or correlation is considered to be sufficient for representation (*cf.* §3.2, this chapter). Aside from making the concept of representation itself almost vacuous, Ramsey concludes that deflationary accounts of representation can in fact hinder our understanding of cognitive systems and so should be rejected.<sup>86</sup>

In short, although Orlandi does conclude that the winning hypothesis should be understood in terms of representation, I have argued that in the context of RSE this conclusion is not warranted. On RSE, the winning hypothesis is not concerned with distal events and it is not de-coupleable. Although it does play the role of guiding and executing action, concluding that it deserves a representational status on this basis alone is misguided, because reliant on acceptance of RDT.

At this point, we have arrived at a non-representational version of PP. “Prediction-signal”, “error-signal”, “prior”, “likelihood”, and “posterior probability” have all been described in non-representational terms. Importantly, no explanatorily beneficial insight was lost by describing these concepts in non-representational terms. Having explained how the key PP posits can be understood in non-representational terms, I will now explain how this non-representational account can be subsumed within RSE and argue that an RSE explanation of rivalry improves upon, and is preferable to, the PP explanation proffered by Hohwy *et al.* If this argument is successful, I will have shown both: that RSE can account for

---

<sup>86</sup> Ramsey argues for this point at length in his *Representation Reconsidered* [2009]. A proto-typical example of it can be found in Freeman & Skarda [1990]. In this paper, Walter Freeman and Christine Skarda explain how their research on the neuroscience of olfaction was actually de-railed because of their acceptance of a representational theory of mentality. Once they eschewed this metaphor, and instead looked at the brain as a non-representational and self-organising dynamical system, their understanding of olfaction developed substantially.

the brain's role in conscious perception; and, that RSE provides a better account of the brain's role in conscious perception than cognitivist theories.

#### **§4: RSE & Rivalry**

Consider the following general over-view of rivalry, which fits quite closely with what would be expected on a biased-affordance-competition framework:<sup>87</sup>

Although coherent rivalry may be found only with a small restricted zone of the visual field, many observations suggest that some form of rivalry interactions can occur across neighbouring zones. First, transitions in dominance from one stimulus to the other often appear like a wave: one stimulus appears to sweep over the other, erasing the latter from visual awareness. This implies that the zones of rivalry are interconnected. Second, even a relatively large stimulus, while susceptible to piecemeal rivalry, will achieve complete dominance significantly more often than would be expected based on independent zones. Third, multiple small rival targets scattered throughout the visual field can engage in synchronized alternations, such that all targets of a given configuration are dominant simultaneously. Fourth, visual features located outside the boundaries of a rival target can influence the predominance of that target, implying that the target's strength is being modulated by its surrounding context. Thus, rivalry includes local competitive interactions, leading to exclusive dominance within a region or zone, as well as more global

---

<sup>87</sup> Cf. Hohwy *et al.* [2008, p. 691] who explicitly argue that biased-competition frameworks cohere extremely well with PP accounts of rivalry.

interactions that can facilitate the spread of dominance across large portions of the visual field. [Blake & Tong, 2008]

Biased-affordance-competition takes environmental affordances to cause competitions between various neuronal states, with the winning affordance being the one which controls an organism's behaviour. On RSE, the winning affordance is the one which instantiates Cognitive Unison and results in 'attentive perception'. Consequently, the RSE account would predict that "rivalry includes local competitive interactions, leading to exclusive dominance within a region or zone, as well as more global interactions that can facilitate the spread of dominance across large portions of the visual field".

Consider the example of rivalry discussed in sections one and two. On the RSE framework, the three potential PP hypotheses (face, house, and face-house) should instead be described as three affordances competing for neuronal dominance. The reason that faces and houses are usually perceived by subjects is itself explained via neuronal biases— human beings are exposed to faces and houses so often in everyday life that their brains possess connections which are skewed toward configurations that support the domination of face and house affordances (for empirical accounts as to how these biasing processes arise and are maintained, see [Anderson, 2014; Anderson & Finlay, 2014; Orlandi, 2014; 2015]). These biases will result in the brain naturally configuring itself in accord with the face or house affordance, but not in accord with a face-house affordance. One could then explain the alteration between face and house domination in terms of the accumulation of error-detection (where error-detection is understood in terms of non-representational causal correlation, as argued for in section three). On a non-representational reading, this would

involve increased neural processing in neural areas responding to the non-dominant affordance which has the upstream consequence of changing the winning affordance.

Indeed, combining PP and biased-affordance-competition in this manner can actually result in a better account of brain-based processing [Clark, 2016, ch. 5]. Both frameworks accept a view of the brain's cognitive architecture as essentially action-oriented and non-modular. Andy Clark argues that PP can be reconciled with, and used to improve upon, Anderson's framework because the concept of "precision-weighting" can explain how brains can configure and re-configure TALoNS ("transiently-assembled-local-neuronal-sub-systems", which were introduced and explained in chapter two of this thesis) on extremely rapid time-scales.

The concept of "precision-weighting" is introduced by PP theorists to explain how the predictive brain can effectively reduce prediction error. We live in an inherently unstable and unpredictable environment and so it is inevitable that there will be errors at all levels of the brain's predictive hierarchy. In order to make the organisation of error-signals manageable the brain requires a mechanism for determining which errors should be closely attenuated and which should be ignored. PP theorists suggest that the brain performs this task by assigning precision-weighting to its error-signals. If an error-signal is accorded high-precision, then it will be allowed to propagate up the cortical hierarchy and as such will have a bigger influence on the creation and modification of the brain's winning hypothesis. Analogously, the lower the precision accorded to an error-signal the smaller the impact it will have on the formation of the brain's winning hypothesis. Precision-weighting is therefore taken to be the instrument used by the brain to make its task of minimising prediction-error manageable.

Clark argues that the concept of “precision-weighting” can be used to explain how TALoNS are formed and dissolved. Precision-weighting determines how much influence a given neural area will have on another given neural area because signals with high-precision will be allowed to propagate throughout the cortical system whereas signals with low-precision will not. Precision-weighting assignments are themselves constantly in flux, and so the amount of weight accorded to a given signal can change rapidly. Precision-weighting therefore allows for the rapid creation and dissolution of neuronal coalitions. Consequently, it can be posited as a potential mechanism through which TALoNS are formed and dissolved, and so applying this PP concept to Anderson’s framework improves it.<sup>88</sup>

In short, one can apply a non-representational PP explanation to rivalry and then accept this account within RSE (via biased-affordance-competition). I can see two reasons to prefer this RSE account of rivalry over cognitivist rivals: (1) it fits better with first-person phenomenological experience; and, (2) it provides a parsimonious account of the relation between the sub-personal, personal, and conscious aspects of perception.

#### ***§4.1: Benefit One– Phenomenological Plausibility***

The first benefit of accepting an RSE account of rivalry is that, by positing biased-affordance-competition as the mechanism which explains rivalry, one arrives at a phenomenologically plausible account of the phenomenon. Affordances are personal level phenomena which are phenomenologically experienced by subjects. It is the human being, for example, that perceives the face as affording certain behaviours. Organisms perceive environmental

---

<sup>88</sup> Of course, “precision-weighting” must itself be explained without invoking representation. I suggest that the concept be understood as a (particularly fast acting, neuro-chemically mediated) non-representational bias.

affordances, not parts of their brains. Rivalry itself, on this framework, can be understood as a competition between two affordances— a “face affordance” and a “house affordance”. This means that the dynamical interactions observed within the brain are caused by the subject themselves perceiving environmental affordances. Consequently, the biased-affordance-competition description of brain processing during rivalry provides one with a description of the effect of personal level affordances on brain-based processing. Thus, brain-based processing is linked-up with phenomenological experience in a phenomenologically intuitive manner [*cf.* §3.2.1., ch. 3, this thesis].

This presents an improvement upon PP, if taken as a stand-alone theory of conscious perception, because that theory faces the (quite familiar) problem of having to provide a phenomenologically compelling explanation of the relation between brain-based processing and personal level conscious experience. Although this is an active area of research within PP, barring a complete conceptual or scientific revolution, it is difficult to see how any particular brain-based account could provide a phenomenologically compelling explanation of the inter-relation between brain-based processing and phenomenological experience. As Susan Hurley and Alva Noë explain:

By contrast, if it is brought to our attention that activity in a certain brain area is correlated with vision, we do indeed still want to ask: “But why does brain activity there go with what it is like to see, rather than to hear or touch?” [Hurley & Noë, 2003, p. 147]

The problem with brain-based accounts is that, regardless of the specifics of the account in question, the identification of a certain brain-based process with a certain experience is always going to appear arbitrary. Furthermore, there is always going to be an air of mystery

surrounding how the phenomenology of conscious perception could be contained within, or identified with, the brain. Even the most enthusiastic proponents of PP admit that this is a problem for the view, and they are quite forth-right in conceding that there has, as yet, been no concrete proposals as to how PP could provide a novel solution to it [Clark, 2013, p. 198-199].

Thus, the first benefit of accepting a non-representational PP account of binocular rivalry and subsuming it within RSE is that one arrives at a phenomenologically compelling account of the relation between brain-based processes and conscious experience. This stands in contrast to PP, taken as a stand-alone account of rivalry, which cannot provide such a phenomenologically compelling explanation.

***§4.2: Benefit Two— Conceptualising and Empirically Explaining the Sub-Personal, Personal, and Conscious Levels of Explanation***

The second benefit of accepting a non-representational account of PP and subsuming it within RSE is that doing so allows one to provide a parsimonious account of the relation between the sub-personal, personal, and conscious levels of explanation. RSE explains personal level perception in terms of behavioural dispositions. The categorical basis of these behavioural dispositions is itself explained to be brain-based— organisms are capable of perceiving only if they possess brain-based sensorimotor knowledge [*cf.* McDowell, 1994]. Brain-based sensorimotor knowledge is understood to be constituted by a series of relations between certain sensory inputs and certain other neural outputs. Given that PP explains brain-based processing to be constituted by sub-personal expectations concerning the



relation between sensory stimulation and movement, it could therefore be used to provide an operationalisation of the relationship between certain sensory inputs and certain other neural outputs at the sub-personal level [*cf.* Seth, 2014]. Consequently, it could be applied to RSE and used to provide a fully worked out empirical explanation of the brain-based, sub-personal aspects of sensorimotor knowledge.

Accepting such an account allows for a parsimonious explanation of the inter-relation between sub-personal, personal, and conscious levels of explanation. On RSE the brain is considered to ‘give rise’ to personal level perception by controlling and driving behaviour. The categorical basis of this perceptual behaviour is the possession and exercise of sub-personal sensorimotor knowledge (which is empirically explained via non-representational PP). Conscious perception is then, similarly, explained entirely in terms of behaviour. On this theory, conscious perception occurs when the perceiving organism’s behavioural interaction with the environment is carried out ‘attentively’, with ‘attentive perception’ empirically implemented in the brain via biased-affordance-competition. Thus, RSE attributes to the brain the role of controlling behaviour, and it explains both perception and conscious perception entirely in terms of behaviour.

There is absolutely nothing mysterious or naturalistically unacceptable about the idea that a brain could control behaviour. Therefore, there is absolutely nothing mysterious about the relation between the sub-personal and personal levels of explanation on RSE. Furthermore, on this framework there is nothing mysterious about the relation between conscious and unconscious personal level perception. The difference between conscious and unconscious perceptual processing is explained to be adverbial in nature, and there is nothing metaphysically suspicious or naturalistically awry with the existence of adverbial

behaviour. Consequently, RSE provides a clear distinction between the sub-personal, personal, and conscious aspects of perception. Moreover, it makes sense of their existence and inter-relation without requiring any leaps of imaginative faith or speculative metaphysical theorising. This stands in stark contrast to PP, which cannot provide a similarly parsimonious account of the inter-relation between these levels of explanation (specifically, because it struggles to account for the conscious level of explanation. See [ch. 3, this thesis]).

In sum, RSE can accept a non-representational version of PP and use it to explain the brain-based aspects of perceptual processing during rivalry. Not only is this non-representational account able to accommodate the same empirical data as a representational version of PP, it can do so whilst providing an overall account of rivalry which is both phenomenologically more plausible and metaphysically more innocuous than cognitivist alternatives.

## **§5: Conclusion**

I began this chapter by outlining the empirical phenomenon known as “binocular rivalry” and explaining why research on this topic is driven by paradigmatically cognitivist concerns. Then, I outlined the PP explanation of rivalry. I next argued that the posits of PP can be understood in non-representational terms, and that invoking representation to explain PP processing is unnecessary because it fails the job-description challenge. Finally, I applied this non-representational account of PP to the case of rivalry, and explained that it could be subsumed within RSE. I argued that the resultant RSE account of rivalry both: is phenomenologically more plausible than its cognitivist rivals; and, that it provides a better

account of the inter-relation between the sub-personal, personal, and conscious levels of explanation.

This chapter was motivated as a response to the objection that theories such as RSE should be rejected because they are incapable of explaining the brain's role in conscious perception. The argument throughout this chapter was that this objection cannot be levelled at RSE, because RSE can accept PP accounts of brain-processing within its framework, and PP is generally thought to satisfactorily account for the brain's role in conscious perception. Thus, I concluded that RSE is capable of explaining the brain's role in conscious perception. Furthermore, having argued that this RSE account has a number of advantages over rival cognitivist accounts (such as a representational version of PP), I concluded that RSE in fact provides a better explanation of the brain's role in conscious perception than rival cognitivist theories.

## Chapter Six: Fictionalism about the Representational Posits of PP

In chapter five I argued that PP could be subsumed within RSE and used to operationalise its brain-based aspects. I explained how one could understand its key posits in non-representational terms, and so concluded that PP can be used to explain the brain's role in conscious perception without violating *knowledge-how* (or, for that matter, *activity*). PP theorists are likely to object, however, that much of the work within PP *requires* representation. Even though we can provide a non-representational construal of some PP explanations, much of the literature on PP simply cannot be made sense of without invoking the concept "representation". The following quote, from Andy Clark, is representative of this general attitude:

Could we perhaps have told our story in entirely non-representational terms? One should always be aware of sweeping assertions about what might, one day, be explanatorily possible! But as things stand, I simply do not see how this is to be achieved. [Clark, 2016, p. 293].

In short, it can be argued that the ascription of representation to PP systems (when taken as a whole) is explanatorily beneficial, and so describing PP systems in terms of representation meets the job-description challenge [Clark, 2015; 2016; Gladziejewski, 2016; Hohwy, 2013, ch. 8; Rescorla, 2015; Seth, 2014].

I am willing to concede that there are some PP explanations which do require the concept of representation. However, I will argue that even when the concept is indispensable we are not required to violate *knowledge-how*. My reason for making this claim is that I think

representation is explanatorily indispensable in some PP explanations for *epistemological*, and not *metaphysical*, reasons. In this chapter I am going to apply a *make-believe approach* to scientific models to PP, and thereby explain how representational posits could be epistemologically indispensable to working scientists and yet metaphysically non-existent. I will conclude that representation in PP can be *epistemologically* indispensable but I will not thereby conclude that in such cases it *metaphysically* exists.

This chapter is structured as follows— in section one, I motivate the adoption of fictionalism about the representational posits of PP. In section two, I outline *make-believe* approaches to fictionalism about scientific models and apply a version of this approach to the case of representation in PP. Finally, in section three, I highlight a number of benefits one arrives at by accepting a fictionalist approach to the representational posits of PP.

### **§1: Motivating Fictionalism**

When concerned with ontological matters, there are (roughly) three attitudes one can take toward the entity in question:<sup>89</sup>

1. Realism: *x* exists.
2. Fictionalism: *x* does not exist. However, it plays a useful explanatory role and so can be treated *as if* it exists.
3. Eliminativism: *x* does not exist.

---

<sup>89</sup> This taxonomy is grossly over-simplified (see Chakravartty, 2010, ch. 1, for a more realistic taxonomy). However, I do not take my simplification to be problematic because I intend only to motivate the idea that fictionalism (broadly construed) presents us with a nice middle road between realism and eliminativism (broadly construed).

Fictionalism about a given posit is often adopted by theorists when there is reason to believe the posit is ontologically suspect, and yet the posit itself is explanatorily indispensable. It is, for example, a popular position within debates about the metaphysics of modality and mathematics. In the case of mathematics, although it is difficult to envisage everyday human practice that does not make reference to mathematical objects, mathematical objects themselves are ontologically suspect (because they do not appear to exist in the same manner as spatio-temporal objects). Adopting fictionalism about mathematical entities enables one to keep them within one's ontology whilst avoiding having to explain how they could causally interact with physical objects.

We have seen that RSE (a non-representational framework) can be improved by making extensive use of work carried out under the PP banner. However, it is arguable that much work within PP makes indispensable use of representational content [Clark, 2015; 2016; Gladziejewski, 2016; Hohwy, 2013, ch. 8; Rescorla, 2015; Seth, 2014]. Accepting representation within RSE would require violating the *knowledge-how* tenet, and violating this tenet would mean that RSE would have to face the problems specific to cognitivism described throughout the thesis. Fictionalism is adopted when theorists both: make indispensable use of a concept; and, the concept is metaphysically problematic. PP explanations may sometimes make indispensable use of representation. However, representation is problematic for a number of reasons (at least, in the context of RSE). Consequently, within the RSE framework, fictionalism about the representational posits of PP is well-motivated.<sup>90</sup>

---

<sup>90</sup>Henceforth, when I mention or discuss the representational posits of PP, I should be understood to be referring to PP accounts which make indispensable use of representation. If a PP account can be straightforwardly accepted without representation, there will obviously be no need to provide a fictionalist account of its representational posits (because there are none).

## **§2: Models As Make-Believe and Predictive Processing**

A number of theorists have recently applied Kendall Walton's theory of pretence to the case of fictionalism about scientific models. They have argued that many scientific models provide theorists with prescriptions to imagine certain scenarios, and that the results of these imaginings allow for predictive and explanatory utility without committing theorists to the literal existence of each posit in their model. In this section I introduce and motivate this approach by outlining Adam Toon's work on the topic. Then I draw on the work of Arnon Levy, and explain how the application of representations to the brain (at least in the case of predictive processing) can be shown to be nothing more than metaphorical talk which is used by neuroscientists as an instrumentally useful tool. I conclude that the indispensability of representation for PP models does not require commitment to the literal existence of representations. It is wholly compatible with the practice of PP cognitive science that cognitive scientists are using representational talk only as a metaphorical tool which allows them to keep track of real causal interactions within the brain.

### ***§2.1: Mimesis as Make-Believe and Scientific Representation***

Kendall Walton's [1990] 'mimesis as make-believe' account of fiction<sup>91</sup> has in recent times been applied to work on scientific models by a number of philosophers [Frigg, 2010; Levy, 2011; Toon, 2012]. According to Walton, works of fiction act as props in games of make-believe which prescribe certain imaginings we must undertake. A number of philosophers of

---

<sup>91</sup> Note, applying Walton's account to scientific models does not require taking a stance on the nature of fiction itself— "I do not want to defend Walton's characterisation of fiction here. As I see it, we may remain neutral on this point. As a result, the make-believe view does not commit us to the claim that models are works of fiction." [Toon, 2012, p. 70].

science have argued that scientific models should similarly be considered props in scientists' imaginative games of make-believe. These models serve to prescribe imaginings the scientist should make with regard to a given system, and they are thought to play an instrumental role in helping scientists understand and predict real causal changes in the world. This means that the models themselves do not accurately represent the world and that they are not supposed to. Rather, they are used by the scientist as aides that help her to understand what is actually happening in the world.

Walton argues that fictions should be seen as analogous to the sorts of make-believe games which are played by children. These games are often based on what Walton labels "generative principles", which are the rules upon which the imaginative game is based. Consider the example of a child playing a game with a bubble machine in their garden. In this game the bubbles are imagined to be space-ships which have escaped from the Death Star, and they must make their way home without being destroyed. Given the rules of this imaginative game, a number of generative principles follow:

1. The bubble machine counts as the Death Star.
2. The bubbles generated count as escaping space-ships.
3. The limits of the back garden, signified by the fence, count as the limits of the Death Star's reach.

From these principles a number of sub-principles obviously follow. For example, a corollary of (2) is that:

1. The larger the bubble, the bigger the spaceship is.
2. Bubbles which pop count as destroyed.



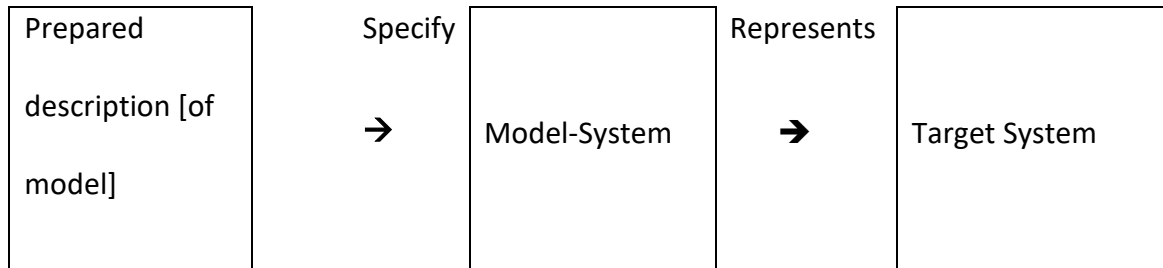
3. Bubbles which float over the garden fence count as having escaped.

Although the child is engaged in a game of imaginative make-believe, the things which happen in the game depend entirely on things which happen in the actual world. For example, the fate of a given space-ship depends entirely on the fate of a bubble in the actual world. Its fate cannot be decided by the whims of the child. Consequently, the child is able to accurately track events which occur in the real world. For instance, they can track how many bubbles float past the garden fence.

### ***§2.2: Why Accept This Model?***

Although a number of theorists have applied Walton's account to scientific models [Frigg, 2010; Levy, 2011] I will here focus on the work of Adam Toon [2012]. Toon differentiates between two different accounts of how modelling works— "*indirect*" and "*direct*". Indirect models are relational and are generally thought to require that scientists learn about the world by representing it via a fictional model.

Figure Twelve: 'Indirect Model'



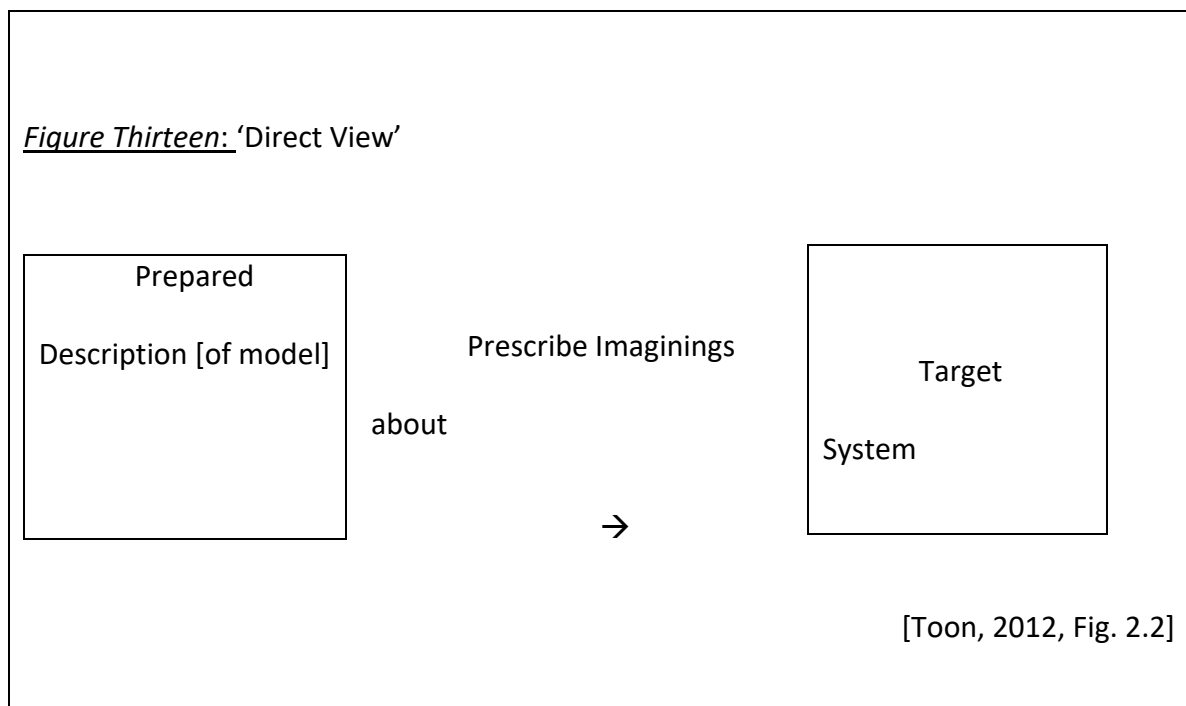
[Toon, 2012, Fig. 1.1]

This model represents when it gets things right and misrepresents when it gets them wrong.

A big problem with the indirect model is that it leads to difficult ontological problems.

Because the idealised model does not literally exist in the world (it describes only an idealised 'model-system') it is hard to make sense of its ontological status. Indirect views often take the status of their models to be akin to the status of fictional posits, which are often thought to exist as abstract objects. This is problematic, however, because fictional posits are notoriously difficult to accommodate within ontological frameworks as abstract objects. For example, consider the fact that abstract objects (such as numbers) are not generally taken to exist in the realm of spatio-temporal objects. Yet fictional entities are supposed to possess spatio-temporal properties (the Death Star is the size of a small moon). As such, it is difficult to accommodate fictional entities under the umbrella of abstract objects because that would entail a contradiction (the Death Star is both abstract and possesses spatio-temporal properties).

Walton's account of fiction was developed specifically with this problem in mind. If we take fictions to prescribe imaginings about their subject matter, then we do not face the problem of accounting for the ontological status of fictional posits (we are not committed to the existence of any such posit). Instead, we are required only to imagine that the posit in question exists.<sup>92</sup> This leads to a direct account of representation— representations represent because they prescribe imaginings. The upshot of this is that the direct account avoids the problem plaguing indirect accounts (that of making sense of fictional entities).



Because Toon accepts Walton's account he argues for a direct view of representation. On this view scientists "learn about a model by exploring the intricate web of imaginings which the model prescribes" [*ibid*, p. 131]. The main advantage of this theory is that it allows us "to explain how it is that models can be representational without representing any real

<sup>92</sup> There lies a similar problem in the vicinity. Rather than having problems accounting for the ontological status of non-existent fictional entities, we are instead faced with the problem of accounting for how we can have contentful imaginings about non-existent objects (such as caloric fluid). I do not discuss this problem here because I am talking about brains which, unlike caloric fluids, do exist.

system” [*ibid.* p. 132]. As will be explained in more detail below, it is chiefly for this reason that I accept the direct view and apply it to the representational posits of PP.

### **§2.3: Levy’s ‘Make-Believe’ Account Of Information In Biology**

Arnon Levy [2011] argues that when biologists make use of the concept of “semantic information” they are engaged in a Walton-esque game of make-believe. Just as the child’s game is constrained by states of affairs in the real world so, according to Levy, the biologist’s use of the concept “information” is constrained by states of affair in the world (at the biological level of description). He outlines three rules, the generative principles of the information-in-biology fiction, which the biologist must (and does) adhere to [*ibid.* pp. 654-55]:

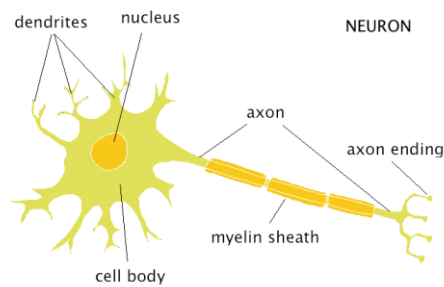
1. *Directionality*: information is transmitted from a sender to a receiver.
2. *Connecting Variation*: transmitted information will cause changes in the receiver (this allows us to gloss over causal events which occur between sender and receiver).
3. *Active vs. passive distinction*: this is a metaphor, which gets at the difference between the parts of the system which change and the parts that do not. The metaphor “active” refers to the changes of state in sender and receiver (of information), whilst the metaphor “passive” refers to the parts of the system which do not change (the information itself).

Levy does note that an argument similar to his own could be given about the use of information talk in neuroscience. However, he does not himself provide such an argument. In the next sub-section I will apply Levy’s three rules to the case of predictive processing.

## ***§2.4: Applying Levy's Fictionalism to Predictive Processing***

I begin my application of Levy's informational principles to the brain by considering the simplest component of the brain (generally) considered relevant for our ability to engage in cognitive activities— the neuron. A neuron is a cell which is composed of three parts: dendrites, soma (or cell body), and axon.

***Figure Fourteen:*** A Neuron



[Boeree, 2009]

Neurons both receive and pass on information in the form of electrical and chemical signals. Electrical and chemical signals are passed across the synaptic cleft (the gap between the axon of one neuron and the dendrite of another) and picked up by the dendrites. These chemical and electrical signals move into the soma of the neuron, and when they reach a certain concentration (the exact concentration can vary from cell to cell) an action potential

will send electrical signals down the axon and cause the neuron to fire. When a neuron fires it passes electrical and chemical signals from its axon to the dendrites of other cells.<sup>93</sup>

At this neurophysiological level of description Levy's three rules are satisfied.

Directionality is satisfied because signals pass from one neuron to another (from the axon of one to the dendrites of another). Connecting variation is also satisfied, because the electrical and chemical constitution of the soma of the receiving neuron will change in response to the electrical and chemical signals received from the neuron which passed on the signal. Finally, the metaphorical distinction between active and passive is also satisfied. Although the electrical and chemical constitution of the two neurons will change, the information being transmitted remains the same. This information is constituted by electrical and chemical signals and these remain unchanged whilst they pass across the synapse (a neurochemical, for example, will not change into another neurochemical whilst passing across the synapse). At this level of description there is no need for the use of representational posits. Indeed, the only predicate in use which is remotely psychological is the term "signal". However, this term refers to electrical and chemical properties and so can be seen as straightforwardly metaphorical. It can be translated into the language of chemistry and physics with no loss in explanatory power.<sup>94</sup>

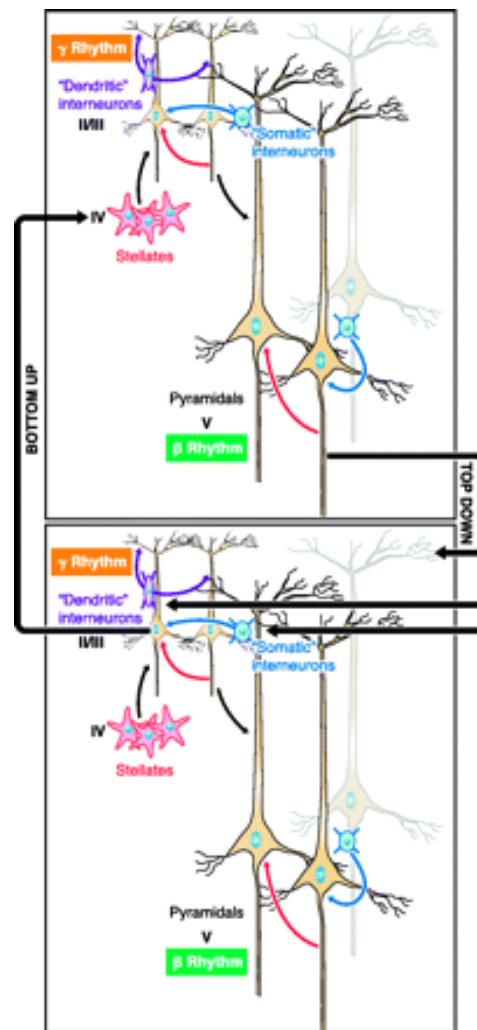
---

<sup>93</sup>This is a simplified description of a typical neuron.

<sup>94</sup> It is arguable that, because the fields of physics and chemistry themselves traffic in informational idioms, an argument for fictionalism in these domains is also required. I leave this as an open avenue for future research.

In order to consider a second model of neural processing, I will now ascend a few levels of description:<sup>95</sup>

*Figure Fifteen:* Top-down signalling between cognitive and sensory areas of the brain



[Wang, 2010, fig. 19]

<sup>95</sup>For an explanation of how to get from a description of a single neuron to a description of interactions between groups of neurons in different parts of the brain, the reader should consult Wang [2010].

This diagram depicts the mechanism through which the passing and regulation of top-down and bottom-up signals occurs between cognitive (top box) and sensory (bottom box) areas of the brain.<sup>96</sup> Neurons located in deep layers of the brain are represented in the lower half of each box. These neurons oscillate in the beta-range and are responsible for sending top-down signals to neurons located in superficial layers of the brain. Neurons in superficial layers are represented in the top half of each box. They oscillate in the gamma-range and send bottom-up signals to neurons in deep layers. The exact nature of the signals sent and received from both deep and superficial layers is dictated by a dynamical inter-play between beta- and gamma- range oscillations [Wang, 2010, fig. 19].

At this higher level of description, which makes reference to groups of neurons in different regions of the brain, Levy's rules are still satisfied. Neurons in the bottom half of each box are designated as the senders of top-down signals whilst neurons in the top half of each box receive these signals. Similarly, neurons in the top half of each box are designated as the senders of bottom-up information whilst neurons in the bottom half are designated as receivers. Consequently, directionality is satisfied. Connection variation is also satisfied because the neurons (whether situated in deep or superficial layers of the brain) will change in response to the signal received. The nature of this change is more complicated than the previous example of the lone neuron. This is because the exact manner in which the neurons change is now dependent on an inter-play between top-down and bottom-up signalling. However, the fact that neurons will change state in response to the signals they receive remains transparent. Finally, the active and passive metaphor is still satisfied. Deep layer neurons are associated with beta-range oscillations, and superficial layer neurons with

---

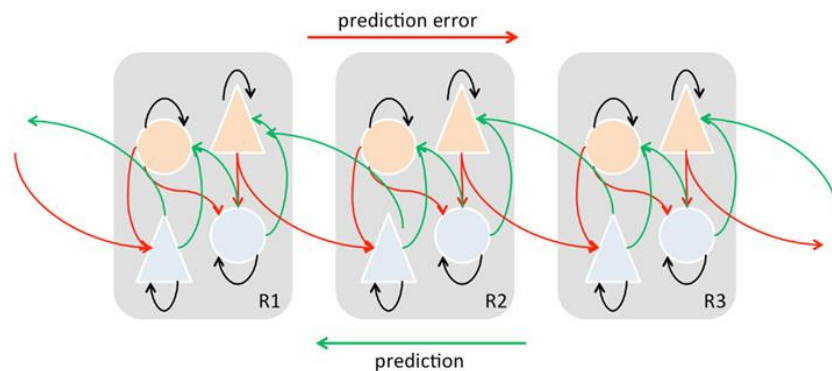
<sup>96</sup>By speaking in this manner I mean only that these areas are causally implicated in cognitive and perceptual processing.



gamma-range oscillations. As such, the oscillations they rely on to transmit information and the type of information they transmit (chemical and electrical signals) never changes, although the states of the neurons themselves will change. Although we have ascended to the level of many neurons which are situated in spatially separated parts of the brain, this description of brain processes is clearly firmly rooted in what we can observe and understand using the tools and framework of neurophysiology. If any psychological predicates are used at this level of explanation it is clear they are being used metaphorically.

I will now consider a third diagram. This diagram is also concerned with top-down and bottom-up neural interactions. However, it is pitched at a slightly higher level of abstraction than the previous diagram.

Figure Sixteen: Predictive processing model of the brain



[Seth, Suzuki, & Critchley, 2012, fig. 3]

This diagram is a representation of how the predictive processing paradigm conceives of the brain. R1 through to R3 represent different levels of the brain, from superficial (R1) to deep

(R3). The green lines are the same top-down signals which are linked with beta-range oscillations that we encountered in figure fifteen. The PP theorists have added to the interpretation made in figure fifteen by labelling the top-down signals prediction-signals and arguing that their role is to transmit predictions down the neural hierarchy. The red lines represent the bottom-up signals associated with gamma-range oscillations that were encountered in figure fifteen. Once more, the PP theorist has added to this interpretation by labelling the bottom-up signals error-signals. On the PP framework these error-signals feedback information that there has been a divergence between the predicted signal and the signal the neuron actually received. Prediction neurons are coloured blue and error neurons are coloured orange [Seth, Suzuki & Critchley, 2012, fig. 3].

Although this model makes use of a lot more representational idioms than any of the previous diagrams, it is obvious that Levy's rules are still being satisfied. The exact same directionality noted in figure fifteen is satisfied because top-down signals (now labelled prediction-signals) still transmit from deep to superficial layers. Similarly, bottom-up signals (now labelled error-signals) display the same directionality, going from superficial to deep layers. Connection variation is still satisfied because the type of signals sent (by neurons in both the deep and superficial layers) will depend on the type of signals they receive. However, the application of psychological predicates allows for a simplification of the exact relation between the top-down and bottom-up signals. By construing top-down and bottom-up signals as prediction- and error- signals respectively, Bayesian mathematics can be applied to the example, and this leads to improved predictions and explanations about the exact nature of the relationship between signalling in the deep and superficial layers of

the brain.<sup>97</sup> Finally, the active/passive metaphor is still in place. The information transmitted stays the same throughout (it will either be a prediction-signal or an error-signal, transmitted in the form of electrical and chemical signals), with only the neurons themselves changing in response to the information received. In short, the exact same causal processes as those in the previous diagrams are being tracked. The only difference is that they are being tracked at a higher level of generality.

Toon defines his 'models as make-believe' [MM] account as follows:

MM: M is a model representation if and only if M functions as a prop in a game of make-believe. [Toon, 2012, p. 62]

I have argued that the PP paradigm adheres to the 'brain-as-representational-Bayesian-machine' game of make-believe. In this particular case the brain is the prop upon which the game of make-believe is predicated (upon which the various 'generative principles' are based). I studied three diagrams of neuronal processing. Each successive diagram represented the brain at a higher level of abstraction, and as the level of abstraction increased so did the use of representational idioms. However, every single diagram satisfied Levy's three principles in the same manner. The exact same causal changes were modelled in each diagram with the only difference being that, as the models became more abstract, the causal changes were modelled in more generality. The neuroscientist is able to ascend from the level of individual neurons, and instead model interactions between groups of neurons (which may be spatially separated to a large degree), by introducing

---

<sup>97</sup>Note, the success of this mathematical explanation of neural activity does not entail the brain is literally engaged in Bayesian calculation. Rather, it entails only that we can use Bayesian mathematics to better understand the causal relationship between top-down and bottom-up signals. Consider, although the orbits of the planets can be calculated using mathematics, it does not follow that the planets themselves are engaged in calculation.

representational predicates. This practice allows the scientist to model real causal processes without having to concern themselves with the minutiae of detail which consideration of the brain as a whole would require. Rather than being seen as pointing to the metaphysical indispensability of brain-based representations, the introduction of representational idioms can be seen as nothing more than a metaphorical and fictional posit which allows the neuroscientist to keep track of real causal changes occurring at the neurobiological level. Treating the brain *as if* it is a representational prediction machine prescribes certain imaginative games, which both help and allow the neuroscientist to track real causal changes occurring within the brain at a greater level of generality.

### **§3: Benefits of Fictionalism**

Thus far, I have motivated a fictionalist approach to the representational posits of PP and explained how such an approach would proceed. In this section, I am going to outline four benefits one accrues by adopting a fictionalist approach to the representational posits of PP.

#### ***§3.1: Benefit One— A Rapprochement of Cognitivist and Enactive/Ecological Cognitive Science***

The most obvious benefit of accepting fictionalism about the representational posits of PP is that one can accept the vast majority of empirical work carried out under the banner of PP within RSE. In chapter five we saw that the posits of PP can be understood without requiring representation. In this chapter, we have seen that even PP explanations which make indispensable use of representation can still be accommodated within RSE (via fictionalism).

Of course, it is an empirical matter whether or not a given PP explanation can be accounted for via non-representational or fictionalist means. However, I submit that adopting both of these strategies would allow one to account for the vast majority of the work carried out on PP and subsume it within RSE.

RSE is advanced within the intellectual tradition of enactive and ecological approaches to conscious perception. Enactive and ecological approaches to mentality tend to respect the tenets of *activity* and *knowledge-how*. As such, they are generally considered to be directly opposed to cognitivist approaches, and so it is often argued that acceptance of one approach requires wholesale rejection of the other [Burge, 2010; Chemero, 2009; Hohwy, 2014; Hutto & Myin, 2013]. RSE does not, however, require a wholesale rejection of cognitivist insights. Quite the contrary, in fact— not only is PP compatible with RSE, it can actually constitute a crucial component of the overall RSE framework (as we have seen). Thus, by accepting RSE and taking non-representational PP<sup>98</sup> to provide an explanation and implementation of the sub-personal aspects of that framework, we arrive at a theory which can take advantage of the empirical work carried out on behalf of both enactive/ecological and cognitivist traditions.

This is beneficial because both of these paradigms have given rise to empirically productive research programmes, which have led to numerous novel and predicted empirical results. Scientific research programmes survive largely on the basis of their empirical productivity, with empirical productivity itself generally thought to require an inference to the best explanation— the theory is empirically productive because it

---

<sup>98</sup>Non-representational PP can be accepted via either: eliminativism about PP's representational posits [ch. 5]; or, fictionalism about PP's representational posits [this chapter].;

accurately describes its domain of study. Although one can reject even empirically successful research programmes (for example, because one thinks they do not correctly describe a given domain of study), in order to do so one must explain how the science can be successful despite presenting us with an incorrect model of the world. Because RSE can accept empirical work carried out in both traditions, it can simply side-step this problem. There is no need to reject, or otherwise eliminate, the vast swathes of empirical work carried out in either scientific tradition. Thus, on RSE, we have no need to reject an inference to the best explanation in either domain of study.

In fact, not only can RSE accept empirical work carried out on behalf of two traditionally opposed scientific frameworks. The framework itself can actually be used to help illuminate the distinction between the two research traditions and provide guidance for future empirical work. RSE typically equates the sub-personal level with the brain,<sup>99</sup> and explains the brain's role to be that of controlling behaviour. Furthermore, it champions non-representational PP as the theory which should be used to study the brain. Thus, if we are interested in investigating the sub-personal aspects of RSE, we can do so by applying the conceptual and empirical tools of non-representational PP to the study of the brain. If, however, we are more interested in investigating the personal level of explanation, then we can do so by using the methods of enactive and ecological approaches to study the interaction, and inter-relation between, the organism and its environment (see, for example, [Chemero, 2009; Varela, Thompson, & Rosch, 1991]). Finally, if we wish to study

---

<sup>99</sup>Note, RSE does not *always* equate the sub-personal level with biological brain-based processing. Rather, it *typically* assumes the brain to be constitutive of sub-personal processing whilst allowing for the possibility that extra-neural factors *could* be playing a constitutive role in such processing [ch. 4, this thesis; cf. Clark & Chalmers, 1998]. Therefore, the equation of the sub-personal level with brain-based processing should be seen as a useful and practical heuristic as opposed to a metaphysical or definitional principle.

conscious perception, we can simply study the manner in which the perceiving organism is able to perceive ‘attentively’ [Anderson, 2014; Mole, 2011]. In short, RSE helps to demarcate between the different levels of explanation within the science of conscious perception, and it provides guidance as to which tools and techniques are appropriate for a given area of study.

### ***§3.2: Benefit Two— Fits with Practise of Cognitive Science***

A second benefit of this fictionalist approach to representation is that it fits with the everyday practice of cognitive science. My account is inspired by, and finds support from, the work of Daniel Dennett [1987]. Dennett argues that representational idioms are used within cognitive science because they provide scientists with instrumentally useful metaphors:

Far from it being a mistake to attribute hemi-demi-proto-quasi-pseudo intentionality to the mereological parts of persons, it is precisely the enabling move that lets us see how on earth to get whole wonderful persons out of brute mechanical parts. That is a devilishly hard thing to imagine, and the poetic license granted by the intentional stance eases the task substantially. [Dennett, 2009, p. 88-89]

Frances Egan argues for a similar claim:

What we normally think of as representational contents— contents defined on distal objects and properties appropriate to the cognitive domain (what I have called ‘cognitive’ contents)— are not in the theory; they are in the explanatory gloss that accompanies the theory, where they are *used* to show that the theory addresses the

phenomena for which we sought an explanation. The gloss allows us to see ourselves as solving problems, exercising rational capacities, occasionally making mistakes, and so on. It characterizes the computational process in ways congruent with our commonsense understanding of ourselves, ways that the theory itself eschews.

[Egan, 2013, p. 131, *italics in original*].

According to Egan, representational explanations in cognitive science provide a user-relative *cognitive gloss* on neuronal mechanisms which helps scientists to understand the workings of the mechanism in question.

The approach to sub-personal representation advocated by both Dennett and Egan is motivated by considerations about the everyday practice of cognitive science. The argument I have provided can be seen to support Dennett and Egan's contention— if PP theorists do use the brain as a prop in the imaginative game of 'brain-as-representational-Bayesian-machine', then PP provides yet another example of cognitive science in which representational posits are in fact being used merely as instrumentally useful metaphors.

Indeed, I think that PP provides a particularly strong example of such an instrumentalist approach to representation. We have already seen that a number of theorists take PP to indispensably require representation only at the more abstract, system-wide level of explanation [Clark, 2015; 2016; Orlandi, 2014; 2015]. This observation would appear to be borne out, more generally, by looking closely at the actual work carried out within PP. In general, representational idioms are most pervasive when PP systems are being described in the abstract. However, when such systems and their implementation are described in greater detail, representational idioms tend to drop out of the resultant explanation. One prominent example, in this regard, is the discussion of the neural



implementation of precision-weighting. Precision-weighting is often described in representational terms:

However, to optimally select the prediction errors...the brain has to estimate or encode their precision. Having done this, prediction errors can then be weighted by their precision, so that only precise information is accumulated and assimilated in high or deep hierarchical levels. [Kanai *et al.*, 2015, p.3]

The use of concepts and phrases such as “estimate”, “encode”, and “informational accumulation” naturally lend themselves to a representation-friendly interpretation. However, when the neural implementation of these representational terms is presented, it becomes clear that what we are discussing here is simply biological neural processing:

This broadcasting of precision-weighted prediction errors may rest on neuromodulatory gain control mechanisms at a synaptic level...This may explain why superficial pyramidal cells have so many synaptic gain control mechanisms such as N-methyl-D-aspartate (NMDA) receptors and classical neuromodulatory receptors like D1 dopamine receptors. [Kanai *et al.*, 2015, p. 3]

Kanai *et al.* do themselves take this neurobiological processing to “correspond[] to a (Bayes-optimal) encoding of precision in terms of the excitability of neuronal populations reporting prediction errors” [*ibid.* p. 3]. However, it is not clear why this reading is required. Certainly, bio-chemical neuronal activity itself does not require a representational explanation. If precision-weighting can be equated with neuromodulation, then there does not appear to be any reason to accept a representational account of precision-weighting. It is thus

typically the case that, when the neural details are specified, a representational explanation is no longer necessary.<sup>100</sup>

I therefore take it to be wholly compatible with the practice of PP theorists that, when representational posits are considered indispensable, they are considered indispensable largely because theorists are engaged in a game of make-believe which treats brains as if they were representing. These games of make-believe are grounded on a number of generative principles which rely on real causal goings on in the brain. Consequently, the metaphors themselves are not relevant to causal interactions in the brain, but rather only to our ability to understand and track these interactions. The practise of PP theorists therefore does not warrant taking the brain to be literally trafficking in representations. Rather, it warrants the weaker claim— that scientists (sometimes) use the brain as a prop in a game of pretence involving representations.

### ***§3.3: Benefit Three— Recommendations for the Direction of Future Research***

Aside from fitting well with the actual practice of theorists working on predictive processing, my account also has direct, practical implications for future research. If I am correct, then it follows that theorists need not concern themselves with finding specific neural implementations of the representational posits of their PP models. These models merely prescribe specific imaginings, and so it is no requirement of them that each representational aspect of the model be found to be isomorphic with some aspect of the world. If my

---

<sup>100</sup> Although this example of precision-weighting has been brief, I think this general pattern (representational idioms are required only when PP systems are described in the abstract) is evident in much of the work on PP. It is particularly evident, for example, in the recent books on PP which have been published by the philosophers Andy Clark [2016] and Jakob Hohwy [2013].

fictionalist account is correct, then it will not be the case that failure to find a neural implementation of (representational aspects of) PP means that it “fails as a distinctive empirical account” [Clark, 2013, p. 200]. As such, theorists can continue to use representational posits in their theories. However, if they do make use of such posits, they should not dedicate their energies toward finding the neural implementation of these representational posits. Earlier in the thesis [ch. 3], we saw that PP theorists tend to take their explanations to require a mechanistic implementation of representational mechanisms. However, acceptance of my fictionalist account requires a turn away from such endeavours. Therefore, fictionalism about the representational posits would have a practical effect on the direction of future research.<sup>101</sup>

### **§3.4: Benefit Four— Sprevak’s IBE**

A common argument against all fictionalist positions is that they are explanatorily unilluminating. In a recent survey concerning fictionalism about sub-personal representation, Mark Sprevak [2013] highlights this as a problem any fictionalist position must face. Normally, when a scientific explanation works, the rule of inference to the best

---

<sup>101</sup>It may be wondered how fictionalism about the representational posits of PP could be compatible with the practise of cognitive science whilst also providing guidance for future research which goes against common practice. To understand why these two conclusions are not incompatible, it helps to introduce the two different types of fictionalism: *hermeneutic fictionalism* and *revolutionary fictionalism*. Hermeneutic fictionalists think that, when engaged in the discourse in question, people already (tacitly) assume that they are speaking in a fictional context. Revolutionary fictionalists argue that, although people who engage in the discourse in question do take themselves to speak literally, we should revolutionise the discourse by encouraging people to stop speaking literally and start taking their discourse to be metaphorical. It is an empirical question which of these approaches to fictionalism correctly describe the practice of theorists working on predictive processing, and in any case I suspect that the answer will vary from person to person (and, indeed, possibly from case to case). In light of this distinction between the two types of fictionalism, I suggest that benefit two be seen as applicable to those who already (if only implicitly) take the representational discourse to be a useful heuristic, whereas benefit three should be taken to apply to those who take the present discourse to require ontological commitment to representation.

explanation [IBE] is invoked. The best explanation for the success of the scientific model is that it accurately describes the phenomenon in question. It seems to follow by IBE that if PP theorists cannot do their work without making reference to “representation”, then sub-personal representations exist. Sprevak notes that this problem is particularly pressing in present circumstances because the fictionalist must “find reasons for rejecting IBE in the case of neural representations that do not apply to other areas of cognitive science where IBE is employed” [2013, p. 557]. My account avoids this IBE problem because in the particular case at issue we have specific, local, reasons for blocking individual attempts (by representationalists) to accept an IBE.

Compare the two competing claims:

- a) IBE: the brain is a representational Bayesian computer
- b) Fictionalism: the brain can be treated *as if* it is a representational Bayesian computer

Speaking explanatorily, (a) and (b) are equivalent because they both allow us to explain the success of representational PP. Theorists tend to accept option (a) because an IBE provides the simplest explanation of why representational PP explanations work. However, as we have seen throughout the thesis, (a) is problematic for theoretical and philosophical reasons (at least, in the context of RSE). I have argued that (b) provides a better explanation of the success of representational PP precisely because it can account for the success of that paradigm *whilst* avoiding the problems realism about sub-personal representation brings.

I am therefore not arguing that we should not use IBE in cognitive science. Nor am I endorsing the claim that we should be sceptical of IBE across the board. Rather, I am claiming that, in this particular instance, we have no need to use IBE. The reason for this is

that a fictionalist explanation of PP representation is better than a literalist one. Applying fictionalism does not require an abandonment of IBE. It instead involves proper application of IBE, because it involves making proper evaluations of what a best explanation actually entails.<sup>102</sup>

#### **§4: Conclusion**

I began this chapter by noting that, although non-representational versions of PP accounts can be provided, some PP accounts make indispensable use of representation. Then, I introduced the position known as “fictionalism”, and explained that a given posit is assigned this status if it is epistemologically indispensable and yet we have metaphysical reasons to doubt its existence. I noted that RSE requires PP in order to explain brain-based processing within the framework, and yet some PP accounts indispensably require representation. Consequently, in the context of RSE, fictionalism about the representational posits of PP was thereby motivated.

The biggest problem facing fictionalist accounts is that of explaining how a posit could be metaphysically non-existent and yet epistemologically indispensable. In order to avoid this problem, one must explain why the posit plays a genuine explanatory role within a scientific theory whilst not actually existing. I applied Levy’s fictionalist account of information in biology to the representational posits of PP, and concluded that these posits do play only an epistemologically indispensable role. By positing representation, the PP theorist is able to abstract away from irrelevant biological details, and this is

---

<sup>102</sup>Of course, more needs to be said on this point. Here, I aim only to motivate the idea that this account gives us good initial (defeasible) reasons to think we avoid the IBE problem.

epistemologically necessary because otherwise they would get lost in a minutiae of biological detail and so be unable to explain the cognitive capacity at issue. I explained how, even when using representational posits, PP theorists are still explaining mere causal neurochemical and biological processes within the brain (none of which themselves involve representation). Thus, I explained how representation could be epistemologically indispensable in PP explanations and yet why it should not be considered to metaphysically exist.

Finally, I highlighted four benefits of accepting this fictionalist approach toward the representational posits of PP. The most obvious benefit is that accepting fictionalism allows for the subsumption of even representational PP within the RSE framework. This allows RSE to accommodate much of the empirical work which has been carried out under the cognitivist and enactive/ecological research traditions (which are generally considered to be incompatible with one another). The second benefit of this account is that it fits with the actual practice of cognitive scientists. The third benefit of the account is that it provides guidance for future research. Finally, this fictionalist account provides (the beginnings of) a response to Sprevak's IBE problem for fictionalist approaches to sub-personal representation.

The upshot of the conclusion of this chapter (and the previous one) is that we have good reason to believe PP can be subsumed within RSE and used to explain its sub-personal, brain-based aspects. Either brain-based processing can be described without invoking representation [ch. 5, this thesis], or, the representational posits in question can be assigned a fictional status [this chapter]. Thus, adopting PP as an account of sub-personal processing

within RSE allows for an explanation of brain-based processing which is compatible with the tenet of *knowledge-how*.

## Chapter Seven: A Note On ‘Non-veridical’ Experience

In this thesis I have developed and defended a direct realist account of conscious perception. Such accounts explain perception to involve a direct and unmediated relation between the perceiving organism and the environmental object being perceived. For example, when visually perceiving a dog, the dog itself is considered to form a constitutive part of my visual experience. I see the dog itself. I do not see a representation or sense-datum which has been caused by the presence of a dog. The question most commonly posed to direct realist accounts goes as follows— direct realism may be capable of accounting for perception, but how does it account for ‘non-veridical’ experiences (such as dreams, hallucinations, and illusions)? In the case of SE, this question has been posed by virtually every cognivist who has grappled with the position in print [Block, 2005; Clark, 2012; Prinz, 2006; Seth, 2014]. Consequently, even though the topic of this thesis is conscious perception (and not non-veridical experience), it seems pertinent to dedicate the final chapter to addressing this popular objection.

Direct realist theories are commonly objected to on the basis of the *arguments from hallucination and illusion*. These arguments proceed in two stages— there is a *base case* and a *spreading step*.<sup>103</sup> The base case consists in making the point that, when we have a non-veridical experience, the thing we take ourselves to perceive is not actually in the environment. As such, this aspect of the experience cannot involve a direct perceptual relation between an organism and its environment. For example, if I have a non-veridical experience of a dog, then I will take myself to perceive a dog even though no such dog is present in the environment. Thus, my experience of a dog in this kind of situation cannot be explained by referring to environmental factors. The spreading step involves taking the base case to be representative of all experience— it is not only the case that *some* types of experience do not involve a direct relation between the experiencing organism and its environment. Rather, *none* of them do. Consequently, experience never involves a direct unmediated relation to the environment, even during instances of veridical perception.

---

<sup>103</sup> As far as I am aware, this terminology was first introduced by Paul Snowdon [1992].



Cognitivist accounts accept the spreading step, and it is in fact for this precise reason that cognitivists spend so much time studying non-veridical experiences. They take these phenomena to provide clues about the particular representational heuristics and strategies our brain employs in order to resolve the problem of perception. Direct realists, however, reject the spreading step. They argue that world-revealing perceptual experiences are experiences of an entirely different type from non-veridical experiences. Consequently, they accept a *disjunctive* approach to experience.<sup>104</sup>

In what follows, I am going to provide a detailed summary of the work of William Fish, who has recently defended direct realism against the arguments from hallucination and illusion [2009]. I will depend heavily on Fish's work in this chapter for two reasons:

1. It is not my aim (in this thesis) to say anything particularly novel or interesting about non-veridical experience. Throughout this thesis I have focused on the topic of conscious perception, and provided a number of arguments for the claim that my own framework can provide a better account of this phenomenon than cognitivist theories. Consequently, I only include Fish's work in order to show that the existence of non-veridical experiences is not incompatible with direct realist accounts. In short, my aim in this chapter is not to provide a distinctive direct realist account of non-veridical experience. Rather, it is to show that such an account can be provided.
2. Disjunctivism about perceptual experience has only existed as an official, properly worked out philosophical position since it was articulated by the philosopher J.M. Hinton [1967; 1973].<sup>105</sup> Since its introduction, disjunctivism has had relatively few proponents,<sup>106</sup> and each of its proponents tend to promote the view for different reasons (which can range from epistemological to phenomenological) and in different ways (ranging from purely conceptual arguments to empirically oriented ones). Of those proponents, the only theorist whose work I find both convincing and

---

<sup>104</sup>Direct realism is not necessarily conceptually equivalent to disjunctivism. However, in practice direct realist theories also tend to be disjunctive theories. In what follows, I will use these terms as synonyms.

<sup>105</sup>Like most positions in philosophy, it is likely that disjunctivism has been advanced in some form or other throughout the history of philosophy. Modern disjunctivism, for example, has antecedents in theorists such as Gilbert Ryle [1949/2000] and J.L. Austin [1962]. However, it is generally accepted that the position in its contemporary form was birthed by J.M. Hinton [Fish, 2009, p. 34-35; Soteriou, 2014, §3.1].

<sup>106</sup>Notable proponents of the view include [Brewer, 2011; Campbell, 2002; Fish, 2009; Martin, 2004; McDowell, 1994; Snowdon, 1992].

consonant with my own aims is that of William Fish. Consequently, this chapter focuses almost exclusively on Fish's disjunctive account.

In short, in this chapter I will echo Fish's claims because: I am only trying to argue for the claim that direct realism (and therefore RSE) should not be rejected solely because non-veridical experiences exist; and, Fish is more-or-less the only theorist who has argued for this point in a manner which I find acceptable.

The chapter itself is structured as follows— in section one I focus on a number of popular conceptual and empirical arguments for the claim that experience supervenes upon the brain. I will summarise Fish's treatment of these issues, and conclude that each argument succeeds only by directly begging-the-question at issue. In section two, I tackle the phenomenon of hallucination and provide sample disjunctive accounts of it. Then, in section three, I outline Fish's direct realist account of illusion. Finally, in section four, I briefly focus on and discuss the extent to which disjunctive accounts can be made to cohere with the *knowledge-how* constraint. I then conclude that, contra the critics, disjunctive accounts can satisfactorily account for the existence of non-veridical experience.

### **§1: Does Experience Supervene Upon the Brain?**

Direct realist accounts take extra-neural factors to be constitutively involved in perceptual experience. Consequently, they require a commitment to externalism about conscious experience. Externalism about consciousness is a controversial position within the mind sciences. It has won few proponents because most scientists and philosophers take it to be scientific fact that experience supervenes upon the brain alone [Block, 2005; Clark, 2012; Koch, 2004; Prinz, 2006; Searle, 1992; Seth, 2014].<sup>107</sup> Direct realism therefore goes against scientific consensus, and so the position is often rejected for being unscientific [Burge, 2005]. William Fish has recently responded to this criticism of disjunctivism. He surveys the four most common arguments provided in support of the thesis that all conscious experience supervenes upon the brain, and shows that each argument begs the question at

---

<sup>107</sup>From now on, when I say "supervenes upon the brain", this phrase should be taken as short-hand for "supervenes upon the brain *alone*".

issue. Instead of proving that experience supervenes upon the brain, the conclusions of these arguments only follow if one assumes that it does [Fish, 2009, ch. 5]. In this section I summarise Fish's arguments, and so show that disjunctivism is not an unscientific position.

### ***§1.1: Argument One: Experience Can Supervene Upon the Brain***

The very existence of conscious experiences like dreaming and hallucination is often thought to require or entail the truth of internalism [Block, 2005; Burge, 2005; Clark, 2012; Prinz, 2006]. Such experiences can occur entirely 'inside the head', and so prove that experience can supervene entirely upon the brain. Given that experience can be constituted by neural activity, it is therefore often assumed that even world-directed perceptual experience is constituted by internal neural processes. Consequently, the existence of dreaming and hallucination is thought to prove that internalism is true.

This argument, in its present form, is quite clearly based on a mistaken inference. The very fact that *some* experiences supervene solely on internal factors does not entail that *all* experiences do [Noë, 2004, p. 211]. Consequently, the mere existence of dreams and hallucinations does not falsify disjunctivism. In order to turn the mere existence of these phenomena into an argument against disjunctivism, one must provide further details. For example, one could use the existence of dreams and hallucination to motivate acceptance of the spreading step, or perhaps explain why disjunctive accounts of dreaming and hallucination are unsatisfactory. I discuss dreaming and hallucination in more detail in section two of this chapter.

### ***§1.2: Argument Two: Brains-in-Vats***

In the brain-in-a-vat thought experiment we are asked to imagine a brain encased in a nutrient-rich vat which is fed the exact same neural inputs which would be received by that brain where it to be embodied in a human being perceptually interacting with the world. It is supposed, by proponents of this thought experiment, that this brain would have experiences identical to the experiences had by embodied beings which are perceptually

interacting with their environment.<sup>108</sup> If a brain-in-a-vat can have the same experiences as an embodied being perceptually interacting with its environment, then perceptual experiences are constituted by brain-based processing. Consequently, the brain-in-a-vat thought experiment proves that internalism is true.

The brain-in-a-vat thought experiment only works as an argument against disjunctivism by explicitly begging the question. It proceeds via the assumption that brains-in-vats would have identical experiences to perceiving beings. This assumption, however, requires the truth of internalism. Consequently, it clearly should not be used as an argument for the position (*cf.* [Clark, 2008; Wilson & Clark, 2008] for arguments to this effect). If the thought experiment is not used in a question-begging manner, then it will no longer work as an argument against disjunctivism. Disjunctivism is not a modal claim to the effect that, necessarily, all conscious perceptual states supervene on something beyond the brain. Rather, it says that in this, the actual world, externalism about conscious perception is true. As such, the mere logical possibility of the brain-in-a-vat scenario does not entail the falsity of disjunctivism. It is only the empirical or metaphysical possibility of a perceptually experiencing brain-in-a-vat which refutes disjunctivism.

At this point in time, it is difficult to envisage how a brain-in-a-vat scenario could be practically implemented [Dennett, 1993, pp. 1-7]. It is even arguable that an empirically feasible brain-in-a-vat would necessarily require the brain to be attached to a body [Thompson & Cosmelli, 2011]. Furthermore, even if we could create an actual brain-in-a-vat scenario, it is difficult to see how one could determine whether or not the brain is supporting perception-like experience without directly begging the question one way or the other [Fish, 2009, p. 126]. Having noted all of the foregoing, it should be evident that perceptually experiencing brains-in-vats are, at present, mere “philosophical fantas[ies]” [Fish, 2009, p. 126]. Philosophical fantasies should not be used as arguments against

---

<sup>108</sup>I have not been able to find examples of theorists who have used this argument explicitly as an argument against direct realism. Fish quotes Michael Tye on this point: “Michael Tye suggests that a range of ‘imaginary cases...play a large role in many philosophers’ thinking’ on this subject (1995: 152). As an example of such an imaginary case, he develops a version of Hilary Putnam’s infamous brain in a vat... Tye is right: thought experiments such as this do seem to play an important role in explaining why many philosophers take local supervenience to be true.” [Fish, 2009, p. 120-121]. Although it is difficult to find published examples of this argument, I do encounter it often enough (in colloquia and the like) to think it warrants discussion.

substantive empirical hypotheses. Therefore, disjunctivism should not be rejected on the basis of brain-in-a-vat thought experiments.

### ***§1.3: Argument Three: Penfield's Experiments***

From the mid-1930s to the mid-1950s, Roger Penfield carried out neurosurgery on over one thousand patients suffering from severe epilepsy. Whilst preparing patients for surgery, Penfield stimulated various parts of their brain with a micro-electrode (in order to ensure he did not remove healthy or important neural areas). These patients often reported experiences such as seeing visions or hearing voices when Penfield stimulated certain parts of their brain. He carefully catalogued these events, noting down when stimulation of a given area neural area resulted in experience, and if so, what kind of experience the stimulation gave rise to. Penfield concluded that stimulation of a patient's brain results in extremely vivid, perception-like experiences [Penfield & Perot, 1963].

Penfield's work is often cited as definitive proof that conscious experience always supervenes upon the brain [Koch, 2004; Prinz, 2006; Lowe, 1992; Robinson, 1994]. It appears to show that perception-like experiences are constituted entirely by brain-based processes. William Fish explains, however, that closer inspection of Penfield's work reveals that the experiences had by subjects are *not at all* similar to perceptual experience. Subsequent investigation of Penfield's own notes, as well as more recent empirical work on neural stimulation, reveals that the patient's reports do not support the claim that their experiences are perception-like. Rather, the patient's reports of their experiences are more accurately described as dream or imagery-like (Fish cites in particular [Mahl *et al.*, 1964; Horowitz *et al.*, 1968]). Fish summarises his point as follows:

[W]hile we can undoubtedly induce visual experiences in some subjects, I suggest the evidence actually points to the conclusion that these visual experiences are not experiences that have phenomenal character that is identical to that a veridical perception would have, as a total local supervenience [thesis that experience supervenes upon the brain] would require. For one, such a claim is difficult, if not impossible, to assimilate with Penfield's own admission that all the patients remain

aware of the operating theater while undergoing these experiences. Second, as some of his own patient reports suggest, and as is borne out in later work, Penfield was overly hasty in assuming that these patients “‘relived’ all that [they] had been aware of in [an] earlier period of time” (1975: 21). [Fish, 2009, p. 133]

It therefore appears that Penfield’s patients did not undergo perception-like experiences. Consequently, Penfield’s experiments show only that some experiences can supervene upon the brain. I have already explained (in §1.2) why the mere fact that some experiences can supervene upon the brain does not show that all experiences do. Therefore, the fact that Penfield was able to induce experiences by stimulating neural areas is not incompatible with, and nor does it require outright rejection of, disjunctive accounts of experience.

#### ***§1.4: Argument Four: NCC Research***

The final popular argument for internalism I will consider refers to work carried out within the NCC research programme (which we encountered in chapter five). To recap, research in this area proceeds by localising a given type of experience to a particular area of the brain and so seeks to find the neural correlates of that experience. This research is thought to show that perceptual experience supervenes upon the brain— certain perceptual experiences have certain neural correlates because the experience in question supervenes upon that neural area [Mormann & Koch, 2007; Koch, 2004; Prinz, 2006]. Consider, for example, that empirical research appears to show the ability to perceive colour to be correlated with activity in area V4 of the brain. If area V4 is ablated or interfered with in certain ways, the subject will be unable to perceive colour. Thus, it is assumed that colour experience supervenes upon area V4 of the brain.<sup>109</sup>

As I explained in chapter five, although NCC research is predicated on the idea that brains alone ‘generate’ conscious experience, the empirical results themselves do not actually require such an assumption. William Fish suggests an alternative interpretation of

---

<sup>109</sup> This is quite a superficial description of what is, in fact, an intricate area of empirical research. However, such a superficial description will suffice for present purposes. My aim here is only to show that this kind of research does not entail the falsity of disjunctivism, and one need not provide a detailed account of NCC research on colour vision in order to achieve this aim.

NCC research which does not require the assumption that brains alone ‘generate’ experience.<sup>110</sup> Rather than having the role of ‘generating’ experience, Fish suggests instead that the function of brains is to enable direct perception of environmental objects:

In order for me to be conscious of a certain feature, I have to (1) have some kind of neural mechanism in my brain that is attuned to register the presence of that feature, and (2) have activity in that neural mechanism. [Fish, 2009, pp. 135-136]

Consider the correlation between activity in V4 and colour experience, which Fish explains as follows: colour experience occurs because colour itself exists as an objective property of the perceived object. The subject is able to perceive colour because they possess the capacity to do so, and this capacity is itself enabled by activity in area V4 of the brain. This alternative conception of NCC research denies that brains alone ‘generate’ experience, and yet it can still accommodate the empirical work carried out within the research programme. This being the case, we can see that the evidence for the existence of neural correlates of experience does not show disjunctivism to be false.

In this section I have provided a summary of Fish’s survey of the four most popular arguments for internalism about conscious experience. Closer scrutiny of each of these arguments reveals that they all fail— the mere existence of some brain-based experience does not prove that all experience supervenes upon the brain, brain-in-vat thought experiments are either question-begging or irrelevant, and NCC research does not require the truth of internalism. In short, Fish shows that none of these popular arguments succeed in refuting disjunctivism. Having said this, in order to motivate acceptance of disjunctivism, a disjunctive account of non-veridical experiences must be provided. Although providing a fully detailed and RSE-specific account of these phenomena goes beyond the scope of this thesis, in what follows I will show that a disjunctive account of these phenomena *can* be provided. Consequently, I will conclude that positions like RSE should not be rejected for being incapable of explaining these phenomena (contra [Burge, 2005; Block, 2005; Clark, 2012; Prinz, 2006; Seth, 2014]).

---

<sup>110</sup>This explanation is quite consonant with the RSE account argued for in chapter five of this thesis.

## **§2: Hallucination**

“Subjectively indiscriminable hallucinations” are experiences in which a subject is having a hallucinatory experience and yet believes themselves to be undergoing a perceptual experience of the world. If such experiences can occur, then their existence would justify acceptance of the spreading step— ‘what it is like’ to experience is the same in both perceptual and hallucinatory cases, and so there is a ‘common-factor’ to all types of experience.<sup>111</sup> In order to provide a satisfactory disjunctive account of experience, a disjunctive account of hallucination must be provided.<sup>112</sup> Disjunctive accounts of hallucination are typically divided into two categories: *negative* and *positive*. In what follows, I will provide examples of both kinds of account. Therefore, I will show that disjunctivist positions can provide a satisfactory account of hallucination.

### **§2.1: Negative Disjunctivism**

*Negative disjunctive* accounts of hallucination explain hallucination via negation. They proceed by explaining hallucination in terms of what it is not. These accounts typically define hallucinations as experiences which appear (to the experiencing subject) to be perceptual in nature, but which are in fact hallucinations. In this section I will outline William Fish’s account of hallucination [2009, ch. 4], because I take it to be the most plausible negative disjunctive account in the contemporary literature.<sup>113</sup>

Fish provides the following definition of hallucination:

For all mental events, *e*, in doxastic setting *D* with cognitive effects *C* (in its subject),  
*e* is a pure hallucination of an *F*, if and only if

- *e* lacks phenomenal character, and

---

<sup>111</sup> It is not clear to me that this assumption— that we can have hallucinatory states which are subjectively indistinguishable from perceptual experience — is in fact justified. I discuss this point in §2.2 of this chapter.

<sup>112</sup> *Mutatis mutandis* for the case of dreaming and (possibly) imagination. I here focus solely on hallucination for brevity’s sake. However, the accounts of hallucination which I will provide can be straightforwardly applied to the cases of dreaming and imagination.

<sup>113</sup> The interested reader can consult [Brewer, 2011; Campbell, 2002; Martin, 2004] for other versions of negative disjunctivism.



- there is some possible veridical visual experience of an  $F$ ,  $V$ , that has a rational subject who is in  $D$  and produces  $C$ , and
- $C$  is nonempty.

[Fish, 2009, p. 94]

In essence, Fish defines a hallucination of  $F$  as a state which has the same cognitive effects that a perceptual experience of  $F$  would have.<sup>114</sup> For example, if a subject is hallucinating a rabbit, they will be led to behave and act in a manner compatible with their having perceived a rabbit. These cognitive effects are then supposed to explain why the subject's hallucination of  $F$  is taken (by that subject) to be indiscriminable from a perceptual experience of  $F$ .

Importantly, however, the hallucination of  $F$  lacks one crucial property which differentiates it from a perception of  $F$ . According to Fish, hallucinations lack *phenomenal character*. Fish has quite a specific definition of phenomenal character in mind:

[T]he phenomenal character of a visual experience is the property of acquainting the subject with a selection of the facts that inhabit the tract of the environment the subject perceives. The particular array of facts that the experience relates the subject to—and hence the phenomenal character of the experience itself—is determined by

- the distribution of objects and properties in the environment,
- the subject's position in/perspective on that environment,

---

<sup>114</sup> Two points are worth noting here: (1) Fish includes the concept of "rationality" in his definition "because we can envisage the possibility of bizarre situations... in which a veridical perception of a certain kind has radically atypical effects." [2009, p. 102] For example, the cognitive effects of a veridical perception of a white van by a person suffering from schizophrenic delusions may result in paranoia about government surveillance, attempts to out-wit and escape the government, and so on. A person who had a hallucination of government agents surveying them from afar may act in a similar manner. If the concept of "rationality" is not included in Fish's definition, we could therefore be led to type this person's hallucination of government agents surveying them as actually being a hallucination of a white van (even though it is a hallucination of government agents, and not white vans). (2) Fish relativizes cognition itself to the cognitive capacities of the hallucinating subject in question. Rats, for example, are unlikely to have hallucinations of government agents surveying them. But they could hallucinate the presence of other rats, and this would then lead to the kinds of cognitive effects and behaviour which usually occur when rats sense the presence of other rats.

- the nature of vision in general and the idiosyncrasies of the subject's visual system,
- the current distribution of the subject's attentional resources, and
- the subject's conceptual resources.

Taken together, these considerations enable us to make sense of the many different ways in which veridical experiences of the same tract of the environment may differ in phenomenal character. [Fish, 2009, p. 75]

The specifics of this definition are not, however, relevant for present purposes. In the context of the present discussion, phenomenal character can be adequately summarised as concerning 'what it is like' to experience. Thus, for Fish, there is 'something it is like' to perceive but there is not 'something it is like' to hallucinate. This critical difference allows Fish to block the spreading step— there is an important difference between perceptual experiences and hallucinatory experiences because there is 'something it is like' to perceive but not 'something it is like' to hallucinate. Consequently, a disjunctive approach to these experiences is motivated.

There do exist cases wherein a subject may hallucinate and yet be aware that they are hallucinating (Fish labels such hallucinations "*resisted hallucinations*"). Such hallucinations may appear problematic for Fish's account because, in these cases, the hallucination of *F* will not have the same cognitive effects as a perception of *F*. For example, my resisted hallucination of a pixie tea-party will not have the same cognitive effects as my veridical perception of a pixie tea-party. I will know that I am hallucinating a pixie tea-party and will not, therefore, act as though I am perceiving a pixie tea-party. Fish explains that resisted hallucinations are not problematic for his account, however, because a resisted hallucination of *F* will have the same cognitive effects as a perception of *F*. Imagine that I have an extremely realistic perceptual experience of a pixie tea-party at a science exhibit. In this case I will know that I am not really perceiving a pixie tea-party (e.g. because I know I am perceiving holographic projections). Consequently, I will take myself to be having the perceptual experience that it is *as if* I perceive a pixie tea-party. According to Fish, a resisted hallucination of a pixie tea-party should be categorised as a hallucination which has the

same cognitive effects as my perceptual experience *as if* I am perceiving a pixie tea-party. In short, in applying Fish's account to cases of hallucination (such as resisted hallucination), we need to be very careful to specify the perceptual state which has the same cognitive effects as those which occur during a particular hallucination. In this manner, Fish's account should be found to be capable of categorising and explaining the various types of hallucination experienced by subjects.

Fish finishes outlining his negative disjunctive account by noting the following objection:

I imagine it might be alleged that this account will be incomplete until some account of *why* hallucinations come to have the same kind of effects as certain kinds of veridical perception has been provided. [Fish, 2009, p. 114, *italics in original*]

The objection being suggested here is that Fish has not explained *why* hallucinations occur. For example, we need an explanation as to *why* a subject takes themselves to be perceiving a pixie tea-party even though no such tea-party occurs. The most natural suggestion to make here is that the subject takes themselves to perceive a pixie tea-party because they have an experience (of some sort) which leads them to believe they are perceiving pixies. Fish rejects this suggestion, however, because he denies that there is 'something it is like' to experience a hallucination. Consequently, there is no experience occurring from which cognitive effects could originate. Instead, he responds to the objection as follows:

Because a mental event qualifies as a hallucination only inasmuch as it has the same kinds of effects as a certain kind of veridical perception, asking why *hallucinations* have these effects would be akin to asking what bachelors have in common in virtue of which none of them are married. Hallucinations *just are* those events that have the same kinds of effects as, and are therefore indiscriminable from, veridical perceptions of a certain kind. [Fish, 2009, p. 114, *emphasis in original*]

In short, Fish's response is that this objection to his account is conceptually incoherent. On his account, a hallucination of *F* is *by definition* an experience which lacks phenomenal character and yet results in the same cognitive effects as a perception of *F*. Consequently, to argue that his account fails because it cannot explain *why* hallucinations occur requires a

conception of hallucination which his own framework does not recognise. Therefore, Fish avoids this problem because it cannot arise from within his framework.

Fish's negative disjunctive account could be straightforwardly applied to RSE. Hallucinations in RSE could be defined as experiences which do not have a phenomenal character (because they do not involve a direct relation to environmental objects) and yet which have the same cognitive effects of direct perceptual experiences of environmental objects. Consequently, the proponent of RSE could accept Fish's negative disjunctive account and therefore provide an explanation of hallucination.

## ***§2.2: Positive Disjunctive Account of Hallucination***

*Positive disjunctive* accounts provide a positive explanation of hallucination because, rather than defining hallucination in terms of what it is not, these accounts deliver a positive characterisation of hallucinatory experience. Disjunctivists (such as [Brewer, 2011; Campbell, 2002; Fish, 2009; Martin, 2004]) tend to reject/avoid positive accounts because it is feared that provision of such an account will license acceptance of the spreading step— if the disjunctivist does provide a positive account of brain-based hallucinations which allows that they are perception-like, then it may be concluded on this basis that brains alone are sufficient for perception-like experience. The problem apparently posed is that provision of a positive story about 'subjectively indiscriminable hallucinations' will require acceptance of the claim that perceptual experiences can supervene on the brain.

In order to provide a positive disjunctive account which does not enable the spreading step in this manner, one must deny that hallucinations can have the same phenomenal character as world-revealing perception. The positive disjunctivist need not deny that there is 'something it is like' to dream or hallucinate. However, in agreeing that there is 'something it is like' to undergo these experiences, they must explain that this 'something it is like' has an entirely different phenomenal character to perception. Providing such an account requires explaining how subjects could take themselves to be having perceptual experiences though they are in fact hallucinating, without allowing that the phenomenal character of perception and hallucination is the same. In what follows, I

will provide a general outline of such a positive disjunctive account by sketching an RSE specific version of the view.

### ***§2.2.1: Phenomenal Difference***

Most people would agree (at least, pre-theoretically) that there is ‘something it is like’ to dream.<sup>115</sup> I submit, however, that this ‘something it is like’ to dream is not at all similar to the ‘something it is like’ to perceptually experience [*cf.* Noë, 2004, ch. 7]. If one attends closely to the actual experience of dreaming, it becomes obvious that dreaming experiences are not at all like perceptual experiences. When we dream the objects we experience appear amorphous, they appear ‘fuzzier’, and the physical laws which apply to those objects are completely different to those which apply during world-involving perceptual experience. In short, there does appear to be an experiential difference between dream experiences and world-revealing perception. Consequently, a disjunctive account of these experiences would appear to be well-motivated from a phenomenological perspective.<sup>116</sup>

Accepting that perceptual experiences differ phenomenologically from dreams and hallucinations provides motivation for avoiding the spreading step of the argument from hallucination. If perceptual experiences do have a different phenomenal character from dreams and hallucinations, then there is no need to argue that the former are experiences of the same type as the latter.

---

<sup>115</sup> I here switch from the example of hallucination to the example of dreaming because most readers will be intimately familiar with dreaming experience whilst (most likely) being not so familiar with hallucination. However, the points I make apply as much to hallucination (and indeed imagery) as they do to dreaming.

<sup>116</sup> Interestingly, even theorists who argue specifically against enactive disjunctive accounts agree with this point. Andy Clark, for example, notes that: “[T]he kinds of internalist model highlighted above [referring to predictive processing] have the pleasing property of presenting perception, dreaming, and imagination within a single unifying framework while leaving plenty of room for the kinds of typical difference (with respect to detail, richness, stability, and overall coherence) highlighted by Noë and the enactivists.” [Clark, 2012, p. 14]. In a similar vein, Anil Seth writes: “Interestingly, while conceding that sensorimotor theory is indeed challenged by synesthesia, Hurley and Noë do suggest that synesthetic color experiences may be phenomenally distinct from normal color experiences, a line of argument developed here as well.” [Seth, 2014, p. 103].

### §2.2.2: Indiscriminability

A positive disjunctive account will accept that there is ‘something it is like’ to undergo dream or hallucinatory experiences. It will even accept that subjects can undergo such experiences and take them to be indiscriminable from perceptual experience.

Indiscriminability itself, however, will be explained in terms of cognitive deficiencies. The hallucinatory experience will not be considered subjectively indiscriminable from perceptual experience because it has the same phenomenal character as perception. Rather, it will be considered to have a different phenomenal character from perceptual experience, with subjects unable to realise this fact because they possess cognitive deficiencies. It should be quite obvious how such an account of cognitive deficiencies would proceed—schizophrenic patients undergoing a psychotic episode may not realise they are hallucinating because of entrenched delusional beliefs, someone under the influence of recreational hallucinogens may not realise they are hallucinating because these substances substantively alter brain functions and cognitive abilities, and so on. Rather than focusing on these familiar examples, I want to briefly focus on one explanation for cognitive deficiencies in hallucination which accounts based on *activity* are particularly well-placed to explain.

Daniel Dennett points out that:

[O]ne of the endemic features of hallucination reports is that the victim will comment on his or her rather unusual passivity in the face of the hallucination. Hallucinators usually just stand and marvel. Typically, they feel no desire to probe, challenge, or query, and take no steps to interact with the apparitions. It is likely...that this passivity is not an inessential feature of hallucination but a precondition for any moderately detailed and sustained hallucination to occur.  
[1993, p. 9]

Alva Noë makes a similar point:

The closest we can come, in the real world, to producing perfect hallucinations— i.e. hallucinatory states that are *indiscriminable* from corresponding perceptual states— is in the psychological laboratory. But here what is usually required, of the perceiver, is that he or she *not* exercise the full range of perceptual skills (e.g. you are asked to

hold still and fixate on a point, etc.). Hallucination in this setting is the breakdown of our perceptual experience; it is not, so to speak, the revving of its engines. [Noë, 2012, p. 44, *italics in original*]

Given how important the thesis of *activity* is in RSE, it is plausible that hallucinatory experiences which appear indistinguishable from perception appear as such because the active nature of cognition has been restricted. When we cognise, we are usually actively engaging with our environments— our bodies are in constant motion, we move artefacts around in our environments, and so on— and these activities play a constitutive role in our cognitive processes [Anderson, 2014; Barrett, 2011; Clark, 2008]. Cognition, as I have stressed throughout this thesis, is a temporally extended and active process. Given that realistic hallucinations tend to occur only when there is an absence of active motor engagements with the environment (for example, whilst floating in sensory deprivation tanks), it is plausible to suppose that these hallucinations (if mistaken to be perceptual in nature) are not recognised as hallucinations because *activity* cannot be or is not exercised. The subject cannot discover, or does not realise that, their experiential state is not perceptual because they are not engaging in cognitive actions. Consequently, hallucinating subjects do have experiences with an entirely different phenomenal character from world-revealing perceptual states. However, due to various restrictions on the active component of cognition, subjects become incapable of realising that the experience is hallucinatory. In short, hallucinations are discriminable from perceptual experience. However, cognitive malfunctioning occurs because *activity* has been restricted, and this results in the subject becoming incapable of realising that they are hallucinating.

### ***§2.2.3: The Phenomenal Character of Hallucination***

Positive disjunctive accounts accept that there is ‘something it is like’ to dream or hallucinate. Consequently, an explanation of why there is ‘something it is like’ to dream or hallucinate must be forthcoming. Different positive disjunctive theories are going to provide different accounts on this point. I will briefly sketch an RSE specific explanation of the phenomenal character of dreaming and hallucination.

In the case of RSE, dreams and hallucinations can be explained to occur when neural areas are activated in such a manner that sensorimotor knowledge is inappropriately exercised [cf. Block & O'Regan, 2012]. When sensorimotor knowledge is inappropriately exercised, the subject undergoes brain processes which are similar to those which would occur when that subject perceptually interacts with the environment. The result of this brain-based processing is that the subject undergoes experiences *as if* they were perceptually interacting with the environment. Importantly, however, these experiences will not be the same as those which would occur in the case of actual perceptual experience. Perceptual experience (on RSE) constitutively involves the brain, body, and environment, whereas non-veridical experiences are constituted by brain activity alone. Thus, experiences of the latter variety are going to be experientially degraded because they involve only one of the three constitutive aspects of perceptual experience (brain-based processing).<sup>117</sup>

The proponent of the 'common-factor' approach is likely to argue that hallucinations can be mistaken for perceptual experiences because both types of experience have the same phenomenal character. In other words, they are likely to object to my claim that veridical and non-veridical experiences have entirely different phenomenal characters. At this point in the argument, we arrive at an impasse— any evidence I provide of hallucinations which are phenomenally distinguishable from perceptual experience will be argued to be weak or straw-man examples, whereas I am going to claim that evidence provided by common-factor theorists can be explained via cognitive deficiencies. Once the debate reaches this point, it is not clear which side shoulders the burden of proof, and nor is it clear how empirical evidence could be used to decisively arbitrate between the two positions. Because it is not my aim to provide a definitive argument for the truth of positive disjunctivism in this chapter, I will not pursue the issue further here. For present purposes, it suffices that I have shown a positive disjunctive account *could* be provided and that such an account is not obviously a complete non-starter. I do take myself to have succeeded in this aim, and so I leave further exploration of this debate as a topic for future research.

---

<sup>117</sup> Furthermore, these types of experience are parasitic upon the presence of perceptual capacities. On this account, perception is explanatorily prior to non-veridical experience because an organism can only mis-deploy sensorimotor knowledge if it already knows-how to perceive.



In summary— I have motivated a positive disjunctive account of non-veridical experience by arguing, on the basis of phenomenological reflection, that there is an important phenomenological difference between world-revealing perception and non-veridical experiences. Consequently, we have no reason to accept the spreading step in the argument from hallucination, because the kind of experience involved in non-veridical experience is that of an entirely different type to that involved in perception. I then argued that subjects can be incapable of discriminating between non-veridical and perceptual experience, even though the difference between the two types of experience is discriminable, because they possess cognitive deficits. I noted that this latter point fits extremely well with accounts that emphasis *activity*, because ‘subjectively indiscriminable hallucinations’ typically occur when subjects do not engage in action. If cognition is essentially active, it is to be expected that restricting action will result in cognitive deficits. Finally, I noted that the very existence of non-veridical experiences can be explained to occur because the brain-processes which cause such experiences are similar to, though not identical with, those which occur during world-revealing perception.

Direct realist accounts are often rejected because they are thought to be incapable of explaining hallucination. I outlined both a negative and a positive disjunctive account of hallucination and so showed that direct realists can provide accounts of hallucination. Therefore, I conclude that direct realism should not be rejected on the basis of the argument from hallucination.

### **§3: Illusion**

The existence of illusory experience is thought to be problematic for direct realist accounts for exactly the same reason that hallucinations are. To recap— direct realism requires that we directly perceive environmental objects. During illusions, we perceive properties or objects which do not exist in the environment. Therefore, the existence of illusions disproves direct realism. In order to avoid the conclusion of this argument I must provide a direct realist account of illusions. In what follows, I outline William Fish’s direct realist account of illusion. Fish distinguishes between three different types of illusion: *physical*, *cognitive*, and *optical*. I will focus on each illusion in turn, and explain how Fish shows the

existence of each to be entirely compatible with the truth of direct realism. Therefore, I will conclude that the existence of illusions does not prove direct realism to be false.

### **§3.1: Physical Illusion**

The first category of perceptual illusion which Fish focuses on are those which he labels “*physical illusions*”. Physical illusions are illusions which can be explained entirely with reference to aspects of the environment. Examples of such illusions include perceiving circular objects to be elliptical, perceiving red objects to be brown, and perceiving straight sticks in water to be bent. Fish argues that physical illusions are not, strictly speaking, illusions. Rather, he claims that they are a special type of veridical perception. He differentiates between two types of physical illusion: shape and colour illusions, and lens illusions. In the next two sub-sections, I will explain how Fish accounts for these phenomena in a manner compatible with direct realism.

#### **§3.1.1: Shape Illusions<sup>118</sup>**

Shape illusions occur when one perceives an object to be shaped a certain way and yet the object in fact possesses an entirely different shape. Consider for example that, from certain viewpoints, a circular coin can appear elliptical. The existence of shape illusions can *prima-facie* appear to provide evidence against direct realism, because we perceive the coin to possess a property (elliptical-ness) which it does not in fact possess (it is circular).

Consequently, it appears that we are not directly perceiving the object in question because the property being perceived is non-existent. Fish argues that, by differentiating between the object’s intrinsic shape and its relational shape, we can see why the existence of shape illusions is not incompatible with direct realism. He argues for this point by using an analogy with mass and weight.

---

<sup>118</sup> In what follows, I focus on the example of shape illusions. However, exactly the same argument can be supplied in the case of colour illusions (wherein one perceives an object to possess a colour which is different from its actual colour). I focus on shape purely for brevity’s sake, but the interested reader can consult [Fish, 2009, ch. 6; Noë, 2004, ch. 4] for an in depth explanation of how colour illusions can be accounted for in exactly the same manner as shape illusions.

Mass is a non-relational property which stays constant regardless of the environmental circumstances of the object in question. Weight, however, is a relational property because an object's weight will vary in different environments due to the effects of gravity. For example, although the mass of an astronaut's boots will stay constant for the duration of the space voyage, their weight will vary depending on their particular location in space. Consequently, although both mass and weight are objective properties of an object, they are objective properties of different types— mass is an intrinsic property whereas weight is a relational property.

According to Fish, when we recognise this distinction between intrinsic and relational properties, we will come to realise that physical illusions do not require the falsity of direct realism. Consider again the example of coins. Coins are intrinsically circular. However, the shape a coin will be perceived to have varies dependent on the relation the perceiving subject has to that coin— if a coin is placed directly in front of my face it will visually appear circular. However, if that same coin is moved further away from and below my face, it will eventually come to appear elliptical. The different ways the coin can appear to me, dependent on our respective positions in physical space, are its relational properties. In an exactly analogous manner to mass and weight, the intrinsic shape of a coin can remain the same whilst being viewed even though its relational appearance properties can change. Thus, when we perceive a coin to be elliptical from certain perspectives, we are still veridically perceiving that coin because we veridically perceive its *relational* property [Fish, 2009, ch. 6; *cf.* Austin, 1962; Campbell & Cassam, 2014; Noë, 2004, ch. 4].

Direct realism is a relational view of experience. Consequently, a relational account of physical illusion is entirely plausible and well-motivated within this framework. By making a distinction between intrinsic and relational shape properties, relational theories can accommodate the existence of shape illusions. Therefore, the existence of such illusions should not be used as a reason to reject direct realism.

### §3.1.2: *Lens Illusions*

Light rays bouncing off objects can become distorted if they pass through certain mediums (such as distorting lenses or water). In such cases, we can come to perceive objects to possess properties which they do not in fact possess. For example, if a straight stick is placed in a tumbler of water it can appear bent. Fish labels these sorts of illusions “lens illusions” (largely because John Foster used the example of lenses in order to provide his own argument from illusion). As was the case with shape illusions, Fish argues that lens illusions are in fact a type of veridical perception.

Fish begins his argument by focusing on how microscopes work. We can use microscopes to observe things which are too small to be observed by the naked eye. They can be used to magnify a certain aspect of *x*, and therefore bring heretofore unobserved parts of *x* into view. For example, if a microscope is focused on a living cell we can come to perceive parts of that cell (such as its structure) which cannot be observed with the naked eye. Fish argues that, in such cases, we come to be appraised of different *facts* about the same object.<sup>119</sup> When we view a cell through a microscope we come to perceive some facts (such as the cell’s structure) at the expense of other facts (the cell is located within a drop of blood). We can remove ourselves from the microscope and so come to perceive certain facts (the drop of blood) but by doing so we are no longer acquainted with certain other facts (the cell’s structure). In each case, we are still directly perceiving aspects of our environment. However, the exact aspects being perceived will depend on the nature of the perceiving subject’s relation to a given aspect of the environment.

Fish then goes on to argue that lens illusions can be explained in an analogous fashion. When we perceive a stick to be bent in water, for example, we come to perceive certain facts about the stick at the expense of certain other facts:

---

<sup>119</sup> Fish provides a detailed account of facts in chapter three of his book. However, space does not permit my expanding on his account. For present purposes, I wish only to note two points: (1) the invocation of facts is used to explain how and why the truth of direct realism is compatible with the existence of the different ‘levels’ of reality described by the various natural sciences; (2) discussion of facts is not taken to require or entail a representational view of perception. When he uses the term “fact”, Fish is simply referring to the different things perceiving subjects can be acquainted with when they are perceptually related to the world.

The explanation of our old friend the bent stick in water will also proceed in this way, but with the water acting as the lens. As before, the distorting action of the water alters the array of facts we can see by precluding us from seeing some facts and enabling us to see others. Because the particular way this occurs will be highly dependent on the particular situation, it will not be possible to give the kind of neat graphic account given above of just how this happens. However, merely by attempting to view something through a regular tumbler full of water, you can see for yourself how, for example, you are precluded from seeing the elements of the scene immediately behind the edges of the glass. The suggestion here is that it is action of this kind that also accounts for the phenomenon known as the bent stick in water by precluding us from seeing the fact of the stick's being straight and altering the way in which we are aware of the various facts pertaining to the part of the stick that lies beneath the surface of the water. [Fish, 2009, p. 165]

As was the case with the microscope, viewing an object through the medium of water will make some facts available at the expense of others. When we view the stick in ideal circumstances, it will visually appear straight because straight sticks have the relational property of appearing as such in ideal circumstances. However, when viewed through a tumbler of water, straight sticks will appear bent because they have the relational property of appearing as such in water-viewing circumstances. In short, just as microscopes make some relational facts available at the expense of others, so too does water make some relational facts about an object available at the expense of others. The existence of lens illusions does not therefore count as evidence against direct realism, because such illusions are considered to involve a (special type of) direct perceptual relation to the environment.

### ***§3.2: Cognitive Illusion***

*Cognitive illusions* occur when subjects directly perceive environmental objects but misinterpret the nature of their perceptual relation to the world. William Fish argues that cognitive illusions occur when subjects mis-deploy their conceptual capacities.<sup>120</sup> This mis-

---

<sup>120</sup>Fish's account of cognitive illusion requires that only organisms with conceptual capacities will be capable of undergoing such illusions. His account of conceptual capacities, however, allows for the ascription of

application of concepts is thought to both result in and explain the subject's subsequent behaviour. Such experiences should be classed as illusions, and not hallucinations, because they still involve the direct perception of environmental objects. Consider an eager bear-watcher, who directly perceives a heap of logs and yet misinterprets their perceptual experience to be one of a bear. According to Fish, the bear-watcher takes themselves to perceive a bear because their conceptual capacities have mistakenly applied the concept "bear" to a perceptual experience of a heap of logs. The bear-watcher will gasp in surprise, take out their camera, and so forth, because they are under the belief that they have perceived a bear. This belief occurs because the subject's conceptual capacities have been mis-deployed. This is still an example of illusion, and not hallucination, because the bear-watcher directly perceives the heap of logs and so these logs form a constitutive part of the experience. Cognitive illusions involve a misguided cognitive reaction to an organism's direct perceptual relation to the environment. Their existence is not, therefore, incompatible with the truth of direct realism.

### **§3.3: Optical Illusion**

*Optical illusions* are defined by Fish as those illusions which must be explained by referring to both environmental factors and to the perceiving subject's perceptual faculties.<sup>121</sup>

Consequently, his explanation of this class of illusion involves aspects of his accounts of both physical and cognitive illusions.

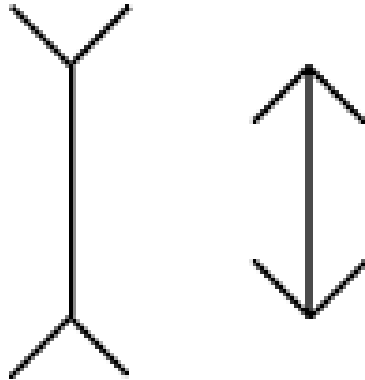
Consider the following well known illusion:

---

concepts to non-linguistics animals [Fish, 2009, ch. 3]. Consequently, he provides quite a permissive definition of concepts. If the reader is unhappy with such a permissive account, they can instead take cognitive illusions to be explained via mis-deployed *cognitive* capacities, and so replace the word "conceptual" with "cognitive" in the above explication of Fish's account. Replacing "concept" with "cognitive" will have no effect whatsoever on the resultant account of cognitive illusion provided by Fish.

<sup>121</sup> Note, these illusions are not necessarily visual in nature. Fish labels them "optical" largely because the most popular examples of such illusions (the Müller-Lyer, variants of the Kanizsa illusion, and so forth) happen to be vision based.

Figure Seventeen: Müller-Lyer Illusion



[Müller-Lyer, 2016]

Both of the lines in the above drawing are the same length. However, we perceive them to differ in length. In this particular illusion, even possessing the knowledge that the two lines are the same length will not change the effects of the illusion. The lines still appear to be different lengths, even after we have been informed that they are in fact the same length.

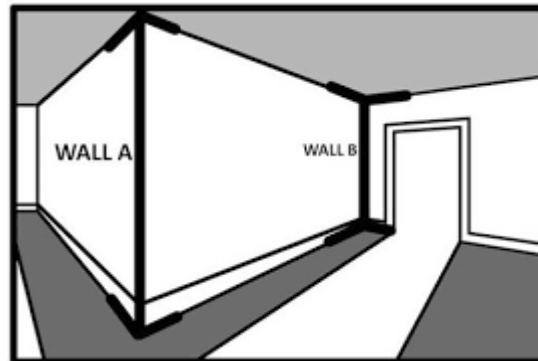
Fish argues, by drawing on the work of Richard Gregory, that the environmental aspect of this illusion can be explained as follows:

[T]he arrowheads therefore provide perspective cues that make the receding corner (arrows out) seem to be farther away than the projecting corner (arrows in). But if the receding corner were indeed farther away than the projecting corner, it would have to be larger in order to cast a horizontal image of the same size on to the retina. This leads us to interpret the receding corner to be larger than the projecting corner, which accounts for the illusory appearance of the arrows-out line as longer.

[Fish, 2009, p. 173]

The explanation being proffered by Fish should become clearer if the reader consults the following diagram:

Figure Eighteen: Perspectival Cues of Müller-Lyer



[Landers, 2014]

Fish's argument is that the Müller-Lyer presents our perceptual system with misleading perspectival cues. These misleading cues trick us into perceiving the visual stimulus to be situated in the environment in a different manner than is actually the case. It thus appears as if the lines are different lengths because usually, when presented with such a visual profile, the object being perceived does have differently shaped relational properties. The mistaken perception of the lines to be different in length then results in a number of cognitive effects, all of which will involve treating the lines to be (or to appear to be) different lengths. Importantly, as was the case with cognitive illusions, this explanation proceeds without denying the direct realist claim that we directly perceive the illusory stimulus itself. The specific manner in which Fish's account is applied to a particular optical illusion will be dependent on the idiosyncrasies of that particular illusion. However, in each case, the explanation is going to proceed in the following manner— first explain the environmental factors which mislead perceptual processing, then explain the particular cognitive effects of this misguided perceptual processing.

In this section I have outlined and endorsed William Fish's direct realist account of illusion. Fish distinguishes between three types of illusion: physical, cognitive, and optical. Physical illusions were explained to involve the direct perception of an object's relational properties. Cognitive illusions were explained to involve the direct perception of an object



which has misguided cognitive effects. Finally, optical illusions were explained to involve the direct perception of stimuli which cue perceptual processing inappropriately and which result in misguided cognitive effects. Consequently, I have shown that direct realists can account for the existence of illusions within their framework.

#### **§4: Non-veridical Experience and Representation**

At this point, the reader may be wondering whether, and to what extent, the accounts of non-veridical experience which have been provided require the concept of “representation”. I want to finish this chapter by briefly sketching out the theoretical landscape of this area, with specific reference to my own RSE account.

##### ***§4.1: Dreaming/Hallucination and Representation***

Dreams and hallucinations involve experiences of non-existent environmental properties. Consequently, it is often assumed that this type of experience requires a representational explanation— the dream or hallucination *misrepresents* the nature of our environment. As I see it, there are two ways RSE could go on this point.

The first option involves simply denying that dreams and hallucinations require representation [Bennett & Hacker, 2003, §6.3-6.31; Chemero, 2009, pp. 129-133; Dennett, 1993, pp. 27-30; Hutto, 2015; *cf.* Ramsey, 2015].<sup>122</sup> On this view, the dreamt or hallucinated ‘objects’ should not be considered to literally exist, and nor should the experience itself be described in terms of correctness conditions. For example, if I dream the presence of a blue cow, the cow itself will not be taken to exist anywhere (even as a represented object) and my dream experience of the cow itself should not be described in terms of correctness conditions. Thus, my dream experience of a blue cow does not require a representational explanation.

---

<sup>122</sup> Note, the cited authors focus specifically on the example of imagery and/or mental simulation. However, there is no reason to think that their arguments could not be extended to the cases of dreaming and hallucination.

A second option involves accepting that an explanation of the dream or hallucination does require representation— the experience itself stands-in for a non-existent state of affairs [Anderson, 2014; *cf.* Clark, 1997; 2008]. Even those who accept thorough-going non-representational accounts of mentality agree that mental representation does exist. Where they disagree with the cognitivist is on the pervasiveness of mental representation. Radical approaches to mentality typically take mental representation to be present only in organisms which possess sophisticated linguistic capacities [Hutto, 2007; Hutto & Myin, 2013], whereas cognitivist approaches typically take representation to exist even at the sub-personal level of explanation [Burge, 2010; Fodor, 1975; Hohwy, 2013].

If RSE were to accept that dreaming and hallucination does require a representational explanation, then an account of how mental representation came to exist must be provided. Non-representational theories typically explain representation to be a sophisticated linguistic capacity which originates from socio-cultural practices. Instead of explaining how representational language came to exist, an RSE account (which accepts a representational explanation of dreams and hallucinations) must instead explain how representational dreams and hallucinations came to exist.<sup>123</sup> If this option were to be pursued, it is vital that any resultant representational account does not allow for the existence of unconscious representation (in order to avoid the conceptual coherence of the ‘hard problem’ [ch. 3, this thesis]). Thus, an RSE account which accepts a representational theory of dreaming and hallucination would most likely have to accept a theory of these phenomena which is quite similar to the accounts proposed within the phenomenal intentionality programme (which takes representation to be constitutively conscious).

Regardless of which particular option is pursued by the RSE theorist, the resultant account of dreaming and hallucination would be capable of accepting either of the disjunctive accounts outlined above. Consequently, RSE can accept the existence of dreaming and hallucination without retracting any of the arguments provided in favour of a non-representational account of perception presented in previous chapters.

---

<sup>123</sup> There has been some empirical work carried out in this area. David Foulkes, for example, has argued that dreaming is a sophisticated cognitive capacity which is gradually acquired by individuals who learn to engage in certain socio-cultural practices [1999].

#### ***§4.2: Representation and Illusion***

Physical illusions were argued to be a special kind of veridical perception. Given that perception itself is explained to be non-representational on RSE, it therefore follows that Fish's account of physical illusion does not require acceptance of representation. Fish's accounts of cognitive and optical illusions may require the concept of "representation", but whether they do will depend entirely on how his own account of conceptual capacities is cashed out. If this account is spelled out entirely in terms of cognitive capacities, then neither cognitive nor optical illusions require a representational explanation because cognition itself need not be explained in terms of representation (as I have argued in depth in previous chapters). If, however, one wishes to endorse a stronger sense of conceptual capacities, such that these involve representation, then one must provide an RSE account of the acquisition of these representational capacities which explains how they arise in nature without invoking the 'hard problem of consciousness' (as was the case for dreams and hallucination). Consequently, once more, the accounts of illusion outlined in this chapter can be accepted within the RSE framework without subsequently requiring that we back-track on the arguments presented in previous chapters.

#### **§5: Conclusion**

I began this chapter by noting that direct realist accounts are often rejected on the basis of the arguments from hallucination and illusion. These types of experience do not appear to involve direct perceptual relations to the environment, and it is often thought that direct realist accounts are incapable of explaining these phenomena. In this chapter I summarized Fish's defence of direct realism against such arguments, and so showed that theories like RSE can accommodate the existence of hallucinations and illusions. I concluded, therefore, that direct realism should not be rejected on the basis of the arguments from hallucination and illusion. Of course, there is a lot more to be said on this topic. Here, I have focused only on arguing for the claim that a direct realist account of non-veridical experiences *can* be provided. In order to motivate acceptance of such an account, one must explain why a direct realist account of non-veridical experience is to be preferred over rival common-factor theories. I do, in fact, think such an argument can be provided. However, spatial

constraints do not allow for a more thorough investigation of this point, and so I leave it as an open avenue for future research.

## Conclusion of Thesis

In this thesis I developed and defended a novel approach to conscious perception, which I labelled “radical sensorimotor enactivism”. In chapter one, I defined the two tenets which would guide the development of my theory: *activity* and *knowledge-how*. By developing a theory which respects both of these tenets, I claimed that we would come to an empirical account of conscious perception which is to be preferred over rival cognitivist accounts (which explicitly reject both tenets).

In chapter two, I outlined the sensorimotor enactive theory of conscious perception and argued that, although compatible with *activity*, extant versions of the view are problematic because they are incompatible with *knowledge-how*. Then, I described how the concepts of “sensorimotor knowledge” and “attention” could be explicated without invoking representation. I claimed that sensorimotor knowledge should be explained in terms of a law-like relation between certain neural inputs and certain other neural outputs (at the sub-personal level) and in terms of Ryle’s theory of knowledge-how (at the personal level). Then, I argued that a non-representational version of Chris Mole’s adverbial theory of attention should be applied to sensorimotor enactivism. I therefore arrived at an account of sensorimotor enactivism compatible with the *knowledge-how* constraint, and so at *radical* sensorimotor enactivism (RSE).

In chapter three, I contended that RSE should be preferred over cognitivist theories of conscious perception because it has the conceptual resources to deflate the ‘hard problem of perceptual consciousness’ whereas cognitivist accounts do not. I argued that the ‘hard problem of perceptual consciousness’ is only conceptually coherent within frameworks which deny the tenets of *activity* and *knowledge-how*. I then explained why RSE

avoids this problem, and moreover, noted that it can do so whilst providing a phenomenologically plausible account of conscious perception. Theories of conscious perception which have serious naturalistic ambitions should explain conscious perception without invoking the 'hard problem'. RSE can do this whereas cognitivist accounts cannot. Therefore, I concluded that we have strong reason to prefer RSE as an empirical approach to conscious perception.

In chapter four I focused on the empirical phenomenon of split-brain syndrome. In part one, I explained what split-brain syndrome is and outlined the four most prominent accounts of split-brain syndrome present in the literature. I argued that each of these accounts is unsatisfactory and, moreover, noted that they are accepted in spite of their deficiencies largely because theorists have an a priori commitment to cognitivism. In part two, I explained how RSE can provide a parsimonious account of split-brain syndrome. I argued that RSE can explain why the conscious experiences of subjects are generally unified and yet why splits of consciousness occur during experimental contexts. Furthermore, I explained how this account could be empirically tested. Cognitivist accounts have always struggled to account for the existence of split-brain syndrome. RSE, however, provides an elegant account of it. I therefore concluded that the example of split-brain syndrome provides one with strong reason to prefer RSE over rival cognitivist theories.

The upshot of chapters three and four is that, by rejecting cognitivism and accepting theories which respect *activity* and *knowledge-how*, new empirical options are made available and advances can be made on some recalcitrant problems within the cognitive science of conscious perception. A common objection to theories predicated on *activity* and *knowledge-how*, however, is that these theories are empirically unilluminating. In particular,

it is often argued that such theories cannot do justice to the (undoubtedly key) role played by the brain in conscious perception. In chapters five and six, I argued that this argument cannot be proffered against RSE.

In chapter five I focused on the empirical example of binocular rivalry. I explained what rivalry is and outlined the PP account of rivalry. Then, I argued that PP can be understood without invoking representation and explained how this non-representational account of PP could be subsumed within RSE. PP is generally thought to provide a good account of brain-based processing and therefore I concluded that RSE can account for the brain's role in conscious perception. Furthermore, I then argued that the subsumption of non-representational PP within RSE in fact provides a better account of the brain's role in conscious perception than rival cognitivist theories (such as representational PP). RSE provides a phenomenologically more plausible account of conscious perception, which can provide a better conceptual and empirical account of the inter-relation between the sub-personal, personal, and conscious levels of explanation. I thus concluded, on this basis, that RSE is (once more) to be preferred over cognitivist accounts as a theory of conscious perception.

In chapter six I conceded that, although many PP explanations can be explained in fully non-representational terms, at a certain point the concept does become indispensable. This concession is potentially problematic for RSE, because if PP requires representation then it cannot be subsumed within RSE. I argued, however, that the indispensable representational posits of PP should be accorded the metaphysical status of a fictional entity. Although representation is indispensable in certain PP explanations, I argued that it is indispensable for epistemological (and not metaphysical) reasons. I argued that the concept

of representation plays an indispensable role in PP explanations because it allows for the cognitive scientist to abstract away from the minutiae of biological detail inside the brain, which would otherwise overwhelm their explanation, and provided an example of how this works in practice in the case of PP. I therefore concluded that, if representation in a given PP explanation is indispensable, it can be accorded the status of a fictional posit. Therefore, even representational PP accounts can be considered entirely compatible with RSE.

Finally, in chapter seven, I focused on the objection most commonly levelled against direct realist theories— that they cannot account for non-veridical experience. Drawing heavily on the work of William Fish, I explained why this objection should not be applied against direct realist theories (such as RSE). I explained why it is fallacious to infer that, because some experiences can supervene upon the brain alone, all of them can. Then, I provided accounts of hallucination and illusion which are compatible with the truth of direct realism. Although there is further work to be done here, I concluded that direct realist theories like RSE should not be objected to solely on the basis that non-veridical experiences exist.

I want to finish this conclusion by briefly highlighting future avenues of exploration for RSE. The most obvious area for further exploration is the topic of non-veridical experiences. In order to motivate RSE as a theory of consciousness *tout court* (as opposed to being only a theory conscious perception), a fully developed account of dreams, hallucinations, mental imagery, and illusions must be provided. I intend to develop the RSE-specific strands of chapter seven in future research. Furthermore, an RSE account of bodily feeling must at some point be forth-coming. I think that such an account could be provided by developing PP theories in this area (e.g. [Seth, Suzuki, & Critchley, 2012]) and making



them compatible with RSE. Once non-veridical experiences and bodily experiences are accounted for, we would be led (I submit) to a complete RSE account of consciousness which could then be applied in clinical settings. This completed theory could be used to help us understand and treat various disorders of consciousness (such as schizophrenia, obsessive-compulsive disorder, and so on).

RSE provides a strong conceptual framework for the empirical study of conscious perception. It shows promise of accounting for a vast array of empirical phenomena, and can even provide explanations of phenomena which have proven recalcitrant within the dominant cognitivist frameworks. I therefore conclude, on the basis of the arguments presented in this thesis, that RSE is worthy of further research, development, and critical assessment.

## Bibliography

- Aizawa, K. (2015). *Perception, Action, and Embodiment*. Retrieved from Academia.Edu:  
<http://newark-rutgers.academia.edu/KennethAizawa>
- Anderson, M. (2014). *After Phrenology: Neural Reuse and the Interactive Brain*. Cambridge, MA: MIT Press.
- Anderson, M. (2015). Precis: After Phrenology. *Behavioural and Brain Sciences*, 1-22.
- Anderson, M., & Finlay, B. (2014). Allocating structure to function: the strong links between neuroplasticity and natural selection. *Frontiers in Human Neuroscience*, 7:918.
- Austin, J.L. (1962). *Sense and Sensibilia*. Oxford: Oxford University Press.
- Barrett, L. (2011). *Beyond the Brain: How Body and Environment Shape Animal and Human Minds*. Oxford: Princeton University Press.
- Bayne, T. (2005). Divided Brains & Unified Phenomenology: a review essay on Micheal Tye's "Consciousness and Persons" . *Philosophical Psychology*, 18:4, 495-512.
- Bayne, T. (2010). *The Unity of Consciousness*. Oxford: Oxford University Press.
- Bayne, T. (2013). Agency As A Marker Of Consciousness. In A. Clark, J. Kiverstein, & T. Vierkant, *Decomposing the Will* (ch. 9). Oxford: Oxford University Press.
- Bechtel, W. (2009). Constructing a Philosophy of Science of Cognitive Science. *Topics in Cognitive Science*, 1: 548–569.
- Bennett, M., & Hacker, P. (2003). *Philosophical Foundations of Neuroscience*. Oxford: Wiley-Blackwell.
- Blake, R., & Tong, F. (2008). Binocular Rivalry. *Scholarpedia*, 3(12):1578.

- Block, N. (1995). On a Confusion about a Function of Consciousness. *Behavioral and Brain Sciences*, 18 (2):227—247.
- Block, N. (2001). Behaviourism Revisited. *Behavioral and Brain Sciences*.
- Block, N. (2005). Review of Action in Perception, by Alva Noë. *Journal of Philosophy*, 102(5).
- Block, N., & O'Regan, J. (2012). Discussion of J. Kevin O'Regan's "Why Red Doesn't Sound Like a Bell". *Review of Philosophy and Psychology*, DOI 10.1007/s13164-012-0090-7.
- Boeree, G. (2009). *Neuron*. Retrieved from General Psychology:  
<http://webpace.ship.edu/cgboer/theneuron.html>
- Bogen, J. (1990). Mental Duality in the Anatomically Intact Cerebrum. In C. Trevarthen, *Brain Circuits and Functions of Mind*. Cambridge: Cambridge University Press.
- Bourget, D., & Chalmers, D. (2014). What Do Philosophers Believe? *Philosophical Studies*, 170 (3):465-500 (2014) .
- Brewer, B. (2011). *Perception and Its Objects*. Oxford: Oxford University Press.
- Brooks, R., & Flynn, A. (1989). Fast, Cheap, and Out Of Control: A Robot Invasion Of The Solar System. *Journal of The British Interplanetary Society*, 42: 478-485.
- Burge, T. (2005). Disjunctivism and Perceptual Psychology. *Philosophical Topics*, 33(1):1-78.
- Burge, T. (2010). *Origins of Objectivity*. New York: Oxford University Press.
- Callaway. (1999). Visual Cortex, Cell Types, and Connections In. In R. Wilson, *The MIT Encyclopedia of the Cognitive Sciences* (pp. 867-868). Cambridge, MA: MIT Press.
- Campbell, J. (2002). *Reference and Consciousness*. Oxford: Oxford University Press.

Campbell, J., & Cassam, C. (2014). *Berkeley's Puzzle: What Does Experience Teach*. Oxford: Oxford University Press.

Carruthers, P. (2006). *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*. Oxford: Oxford University Press.

Carruthers, P. (2016). Higher-Order Theories of Consciousness. In E. Zalta, *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/fall2016/entries/consciousness-higher/>.

Carter, J.A., & Pritchard, D. (2015). Knowledge-How and Epistemic Luck. *Noûs*, 49(3): 440-453.

Cavedon-Taylor. (2011). Perceptual Content and Sensorimotor Expectations. *Philosophical Quarterly*, 61(243): 383-391.

Chakravartty, A. (2010). *A Metaphysics for Scientific Realism: Knowing the Unobservable*. Cambridge: Cambridge University Press.

Chalmers, D. (1996). *The Conscious Mind*. New York: Oxford University Press.

Chalmers, D. (2010). *The Character of Consciousness*. Oxford: Oxford University Press.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society of Biological Sciences*, 362(1485):1585–1599.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Back Together*. Cambridge, MA: MIT Press.

Clark, A. (1999). Visual Awareness and Visuomotor Action. *Journal of Consciousness Studies*, 1-18.

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford: Oxford University Press.

Clark, A. (2009). Spreading the joy? Why the machinery of consciousness is (probably) still in the head. *Mind*, 118 (472): 963-993.

Clark, A. (2012). Dreaming the Whole Cat: Generative Models, Predictive Processing, and the Enactivist Conception of Perceptual Experience. *Mind*, doi: 10.1093/mind/fzs106.

Clark, A. (2013). Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. *Behavioural and Brain Sciences*, 36(3):181- 204.

Clark, A. (2015). Radical Predictive Processing. *The Southern Journal of Philosophy*, 53(1):3-27.

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. New York: Oxford University Press.

Clark, A., & Toribio, J. (1994). Doing Without Representing? *Synthese*, 101: 401-431.

Cohen, M., & Dennett, D. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, (8):358-64.

Craver, C., & Tabery, J. (2016). Mechanisms in Science. In E. Zalta, *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/fall2016/entries/science-mechanisms/>.

Dainton, B. (2000). *Stream of Consciousness*. London: Routledge.

Dennett, D. (1969/2010). *Content and Consciousness*. London: Routledge.

Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

Dennett, D. (1993). *Consciousness Explained*. London: Penguin.

Dennett, D. (2009). Philosophy as Naive Anthropology. In M. Bennett, D. Dennett, P. Hacker, J. Searle, & D. Robinson, *Neuroscience and Philosophy: Brain, Mind, and Language* (pp. 73-95). New York: Columbia University Press.

- Descartes, R. (1641). *Meditation VI, Meditations on the First Philosophy*.
- Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, 18:193-222.
- Dretske, F. (1988). *Explaining Behaviour*. Cambridge, MA: MIT Press.
- Dreyfus, H. & Dreyfus, S. (1980). A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition. *Operations Research Centre*.
- Dreyfus, H. (2002). A Phenomenology of Skill Acquisition as the basis for a Merleau-Pontian Non-representationalist Cognitive Science. . *Proceedings of the International Conference 'Foundations and the Ontological Quest'*. Vatican City, Rome.
- Eccles, J. (1965). The Brain and the Unity of Conscious Experience. In *Nineteenth Arthur Stanley Eddington Memorial Lecture*. Cambridge: Cambridge University Press.
- Egan, F. (2013). How To Think About Mental Content. *Philosophical Studies*, 170 (1): 115-135.
- Felleman, D., & Van Essen, D. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cortex*, 1:1-47.
- Ferguson, S., Rayport, M., & Corrie, W. (1985). Neuropsychiatric Observations on Behavioral Consequences of Corpus Callosum Section for Seizure Control. In A. Reeves, & D. Roberts, *Epilepsy and the Corpus Callosum* (pp. 501-514). Springer.
- Fish, W. (2009). *Perception, Hallucination, and Illusion*. Oxford: Oxford University Press.
- Fodor, J. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.
- Fodor, J. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.
- Fodor, J. (2009). Where is my mind? *London Review of Books*, 13-15.

- Foulkes, D. (1999). *Children's Dreaming and the Development of Consciousness*. Cambridge, MA: Harvard University Press.
- Freeman, W., & Skarda, W. (1990). Representations: Who Needs Them? In J. McGaugh, J. Weinberger, & G. Lynch, *Brain Organization and Memory Cells, Systems, & Circuits* (pp. 375-380). New York: Oxford University Press.
- Frigg, R. (2010). Models and fiction. *Synthese*, 172(2):251-268.
- Friston, K. (2012). Embodied inference and spatial cognition. *Cognitive Processing*, 1-9.
- Gazzaniga, M. (1969). Cross-Cuing Mechanisms and Ipsilateral Eye-Hand Control in Split-Brain Monkeys. *Experimental Neurology*, 11-17.
- Gazzaniga, M. (2000). Cerebral specialization and interhemispheric communication: Does the corpus callosum enable the human condition? . *Brain*, 123: 1293-1326.
- Gazzaniga, M. (2012). *Who's In Charge?: Free Will and the Science of the Brain*. Robinson.
- Gazzaniga, M. (2015, September 21). *Early Split-Brain Research- An Interview with Michael Gazzaniga*. Retrieved from youtube.com: <https://www.youtube.com/watch?v=0lmfxQ-HK7Y>
- Gazzaniga, M., & Le Doux, J. (1978). *The Integrated Mind*. New York: Plenum Press.
- Ghosh, A. ([http://www.khaskhabar.com/images/picture\\_image/1397197526-conjoind-twins-10.jpg](http://www.khaskhabar.com/images/picture_image/1397197526-conjoind-twins-10.jpg)). Shivanath and Sharam. India.
- Gibson, J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin.
- Gladziejewski. (2015). Explaining Cognitive Phenomena with Internal Representations: A Mechanistic Perspective. *Studies in Logic, Grammar and Rhetoric*, 40(1):63-90.
- Gladziejewski, P. (2015). Predictive coding and representationalism. *Synthese*, 1-24.

- Gobel, E., Parrish, T., & Reberb, P. (2011). Neural Correlates of Skill Acquisition: Decreased Cortical Activity During a Serial Interception Sequence Learning Task. *Neuroimage*, 58(4): 1150–1157.
- Hinton, J. (1967). Visual Experiences. *Mind*, 76: 217–227.
- Hinton, J. (1973). *Experiences: An Inquiry into Some Ambiguities*. Oxford: Clarendon Press.
- Hohwy, J. (2013). *The Predictive Mind*. Oxford: Oxford University Press.
- Hohwy, J. (2014). The Self-Evidencing Brain. *Noûs*, doi: 10.1111/nous.12062.
- Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, 108(3): 687–701.
- Horowitz, M., Adams, J., & Rutkin, B. (1968). Visual Imagery on Brain Stimulation. *Archives of General Psychiatry*, 19:469–486.
- Hurley, S. (1998). *Consciousness In Action*. Cambridge, MA: Harvard University Press.
- Hurley, S. (2001). Perception and Action: Alternative Views. *Synthese*, 129: 3–40.
- Hurley, S. (2003). Action and the unity of consciousness, and vehicle externalism. In A. Cleeremans, *The Unity of Consciousness: Binding, Integration, and Dissociation* (ch. 3). New York: Oxford University Press.
- Hurley, S., & Noë, A. (2003). Neural Plasticity and Consciousness. *Biology and Philosophy*, 18 (1): 131–168.
- Hutto, D. (2005). Knowing what? Radical versus conservative enactivism. *Phenomenology and the Cognitive Sciences*, 4:389–405.
- Hutto, D. (2007). *Folk Psychological Narratives: The Sociocultural Basis of Understanding Reasons*. Cambridge, MA: MIT Press.



- Hutto, D. (2015). Overly Enactive Imagination? Radically Re-imagining Imagination. *The Southern Journal of Philosophy*, 53 (1): 68-89.
- Hutto, D., & Myin, E. (2013). *Radicalizing Enactivism: Basic Minds Without Content*. Cambridge, MA: MIT Press.
- Johansson, P. (2008). *Choice Blindness Summary*. Retrieved from [http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCAQFjAAAhUKEwix-Pb0x5fIAhXEy4AKHfLaDTI&url=http%3A%2F%2Fwww.lucs.lu.se%2Fwp-content%2Fuploads%2F2011%2F01%2FChoice-Blindness-summary.pdf&usg=AFQjCNFUezlrC5XvVFawp5KlhX\\_eA9D58w](http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCAQFjAAAhUKEwix-Pb0x5fIAhXEy4AKHfLaDTI&url=http%3A%2F%2Fwww.lucs.lu.se%2Fwp-content%2Fuploads%2F2011%2F01%2FChoice-Blindness-summary.pdf&usg=AFQjCNFUezlrC5XvVFawp5KlhX_eA9D58w)
- Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to Detect Mismatches. *Science*, 310: 116-119.
- Johansson, P., Hall, L., Sikström, S., Tärning, B., & Lind, A. (2006). How something can be said about telling more than we know. *Consciousness and Cognition*, 15: 673-692.
- Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1668): 1-13.
- Kingstone, A., & Gazzaniga, M. (1995). Subcortical transfer of higher order information: more illusory than real? *Neuropsychology*, 9:321-8.
- Koch, K. (2004). *The Quest For Consciousness: A Neurobiological Approach*. Englewood, Colorado: Roberts and Company.
- Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Landers, H. (2014). *Size Constancy*. Retrieved from Intro To Psychology: <http://haleysintrotopsychology.weebly.com/>

- Levine, J. (2004). *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.
- Levy, A. (2011). Information in Biology: A Fictionalist Account. *Noûs*, 45(4): 640-657.
- Lockwood, M. (1989). *Mind, Brain and the Quantum*. Oxford: Blackwell Publishers.
- Lockwood, M. (1994). Issues of unity and objectivity. In C. Peacocke, *Objectivity, Simulation, and the Unity of Consciousness* (pp. 89-95). Oxford: Oxford University Press.
- Lowe, E. (1992). Experience and Its Objects. In T. Crane, *The Contents of Experience: Essays on Perception* (pp. 79-104). Cambridge: Cambridge University Press.
- Mahl, G., Rothenberg, A., Delgado, J., & Hamlin, H. (1964). Psychological Responses in the Human to Intracerebral Electrical Stimulation. *Psychosemantic Medicine*, 26:337–368.
- Malcolm, N. (1959). *Dreaming*. London: Routledge.
- Marks, C. (1981). *Commissurotomy, Consciousness, and the Unity of Mind*. Cambridge, MA: Bradford Books: MIT Press.
- Marr, D. (1982). *Vision*. Cambridge, MA: MIT Press.
- Martin, M. (2004). The Limits of Self-Awareness . *Philosophical Studies*, 120: 37-89.
- Martinez-Conde, S., Macknik, S., & Hubel, D. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5: 229-240 .
- McDowell, J. (1994). *Mind and World*. Cambridge, MA: Harvard University Press.
- McDowell, J. (1994). The Content of Perceptual Experience. *The Philosophical Quarterly*, (44)175: 190-205 .

- Milkowski, M. (2016). A Mechanistic Account of Computational Explanation in Cognitive Science and Computational Neuroscience. In M. Müller, *Computing and Philosophy* (pp. 191-205). Cham: Springer International.
- Millikan, R. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
- Mole, C. (2011). *Attention is Cognitive Unison*. London: Oxford University Press.
- Mormann, F., & Koch, C. (2007). Neural correlates of consciousness. *Scholarpedia*, 2(12): 1740.
- Nagel, T. (1971). Brain Bisection and the Unity of Consciousness. *Synthese*, 396-413.
- Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.
- Noë, A. (2012). *Varieties of Presence*. Cambridge, MA: Harvard University Press.
- Noë, A. (2015). Concept Pluralism, Direct Perception, and the Fragility of Presence. *Open-Mind Collection*.
- O'Keefe, J., & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford: Oxford University Press.
- O'Regan, J. (2011). *Why Red Doesn't Sound Like a Bell: Understanding the Feel of Consciousness*. Cambridge, MA: MIT Press.
- O'Regan, J., & Noë, A. (2001). A Sensorimotor Account of Vision and Visual Consciousness. *Behavioral and Brain Sciences*, 24(5):939-1031.
- Orlandi, N. (2014). *The Innocent Eye: Why Vision Is Not A Cognitive Process*. Oxford: Oxford University Press.
- Orlandi, N. (2015). Bayesian Perception is Ecological Perception. *Minds Online*.  
<http://mindsonline.philosophyofbrains.com/2015/session2/bayesian-perception-is-ecological-perception/>.

- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Penfield, W., & Perot, P. (1963). The Brain's Record of Auditory and Visual Experience: A Final Summary and Discussion. *Brain*, 86:595–696.
- Prinz, J. (2006). Putting the Brakes on Enactive Perception. *Psyche*, 12(1).
- Puccetti, R. (1981). The Case for Mental Duality: Evidence From Split-Brain Data and Other Considerations. *Behavioral and Brain Sciences*, 4 (1):93-123.
- Putnam, H. (1975). The Meaning of Meaning. In H. Putnam, *Philosophical Papers, Vol. II : Mind, Language, and Reality*. Cambridge: Cambridge University Press.
- Ramsey, W. (2009). *Representation Reconsidered*. Cambridge: Cambridge University Press.
- Ramsey, W. (2015). Must Cognition Be Representational? *Synthese*, 1-18.
- Rasmussen, T., & Milner, B. (1977). The role of early left-brain injury in determining lateralization of cerebral speech functions . *Annals of the New York Academy of Sciences*, 299:355-69.
- Rescorla, M. (2015). Bayesian Perceptual Psychology. In M. Matthen, *The Oxford Handbook of Philosophy of Perception*. Oxford: Oxford University Press.
- Robinson, H. (1994). *Perception*. London: Routledge.
- Ryle, G. (1949/2000). *The Concept of Mind*. Chicago: Chicago University Press.
- Schechter, E. (2014). Partial Unity of Consciousness: A Preliminary Defense. In D. Bennett, & C. Hill, *Sensory Integration and the Unity of Consciousness* (ch. 15). Cambridge, MA: MIT Press.
- Searle, J. (1992). *The Rediscovery of Mind*. Cambridge, MA: MIT Press.
- Seth, A. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synaesthesia. *Cognitive Neuroscience*.

Seth, A., Suzuki, K., & Critchley, H. (2012). An Interoceptive Predictive Coding Model of Conscious Presence. *Frontiers In Psychology*, 2: 395.

Shaffer, D., & McBeath, M. (2002). Baseball Outfielders Maintain a Linear Optical Trajectory When Tracking Uncatchable Fly Balls. *Journal of Experimental Psychology: Human Perception and Performance*, 28(2): 335-348.

Shapiro, L. (2010). *Embodied Cognition*. London: Routledge.

Skillen, J. (2009, May). *Scottish Sensory Centre*. Retrieved from The Visual Pathway:

<http://www.ssc.education.ed.ac.uk/courses/vi&multi/vmay092ii.html>

Snowdon, P. (1992). How to Interpret 'Direct Perception'. In T. Crane, *The Contents of Experience: Essays on Perception* (ch. 3). Cambridge: Cambridge University Press.

Soteriou, M. (2014). The Disjunctive Theory of Perception. In E. Zalta, *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/sum2016/entries/perception-disjunctive/>.

Sperry, R. (1966). Brain bisection and mechanisms of consciousness. In J. Eccles, *Brain and Conscious Experience* (pp. 298-313). Heidelberg: Springer-Verlag. Reprinted from (1965): Pontifical Academy Seminar on Study of Brain and Consciousness.

Sperry, R. (1968). Hemispheric Disconnection and Unity in Conscious Awareness. *American Psychologist*, 23: 723-33.

Sporns, O. (2010). *Networks of the Brain*. Cambridge, MA: MIT Press.

Sprevak, M. (2013). Fictionalism About Neural Representations. *The Monist*, 96(4): 539-560.

Stanley, J. (2011). *Know-How*. Oxford: Oxford University Press.

Stanley, J., & Williamson, T. (2001). Knowing How. *Journal of Philosophy*, 98 (8):411-444.

- Strawson, P.F. (1979). Perception and Its Objects. In *Perception and Identity: Essays Presented to A.J. Ayer*. MacDonald, G.F. (ed.). London: Macmillan
- Tanney, J. (2015). Gilbert Ryle. In E. Zalta, *Stanford Encyclopedia of Philosophy*.  
<http://plato.stanford.edu/archives/spr2015/entries/ryle/>.
- The Muller-Lyer Illusion*. (2016, September 20). Retrieved from Rochester Institute of Technology:  
<https://www.rit.edu/cla/gssp400/muller/muller.html>
- Thompson, E. (2007). Look again: Phenomenology and mental imagery. *Phenomenology and the Cognitive Sciences*, 6 (1-2): 137-170.
- Thompson, E., & Cosmelli, D. (2011). Brain in a Vat or Body in a World: Brainbound versus Enactive Views of Experience. *Philosophical Topics*, 39:163-180.
- Toon, A. (2012). *Models as Make-Believe: Imagination, Fiction and Scientific Representation*. London: Palgrave-MacMillan.
- Toon, A. (2015). Where is the Understanding? *Synthese*.
- Tovar-Moll, F., Monteiro, M., Andrade, J., Bramati, I., Vianna-Barbosa, R., Marins, T., Rodrigues, E., Dantas, N., Behrens, T.E.J., de Oliveira-Souza, R., Moll, J., & Lent, R. (2014). Structural and functional brain rewiring clarifies preserved interhemispheric transfer in humans born without the corpus callosum. *Proceedings of the National Academy of Sciences of the United States of America*, 111(21): 7843–7848.
- Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of Royal Society B - Biological Sciences*, 353(1373): 1295–1306.
- Trevarthen, C. (1974). Analysis of cerebral activities that generate and regulate consciousness in commissurotomy patients. . In S. Dimond, & J. Beaumont, *Hemisphere Function in the Human Brain*.

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*.

Cambridge, MA: MIT Press.

Tye, M. (2000). *Consciousness, Colour, and Content*. Cambridge, MA: MIT Press.

Tye, M. (2003). *Consciousness and Persons*. Cambridge, MA: MIT Press.

Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human*

*Experience*. Cambridge, MA: MIT Press.

Walton, K. (1990). *Mimesis as Make Believe: On the Foundations of the Representational Arts*.

Cambridge, MA: Harvard University Press.

Wang, X. (2010). Neurophysiological and Computational Principles of Cortical Rhythms in Cognition.

*Physiological Reviews*, 90(3): 1195–1268.

Ward, D. (2012). Enjoying the Spread: Conscious Externalism Reconsidered. *Mind*, (483):731-751.

Ward, D. (2015). Achieving Transparency: An Argument For Enactivism. *Philosophy and*

*Phenomenological Research*, 92(3).

Ward, D. (2016). Hurley's Transcendental Enactivism. *Journal of Consciousness Studies*, 23(5-6):12-

38.

Watzl, S. (2011). Review of "Attention is Cognitive Unison". *Notre Dame Philosophical Review*,

<http://ndpr.nd.edu/news/27026-attention-is-cognitive-unison-an-essay-in-philosophical-psychology/>.

Wilson, R., & Clark, A. (2008). How to Situate Cognition: Letting Nature Take Its Course. In M. Aydede

(ed.), *The Cambridge Handbook of Situated Cognition* (ch. 4). Cambridge: Cambridge

University Press.

Wittgenstein, L. (1953/2009). *Philosophical Investigations*. London: Blackwell.